DOCUMENT RESUME

ED 431 820 TM 030 000

AUTHOR	Glass, Gene V., Ed.
TITLE	Education Policy Analysis Archives, 1999.
ISSN	ISSN-1068-2341
PUB DATE	1999-00-00
NOTE	525p.; "Education Policy Analysis Archives" is an electronic-only journal (see http://epaa.asu.edu/epaa) covered on an article-by-article basis in "Current Index to Journals in Education" (CIJE). For the 17 English-language articles in this volume, see TM 521 898-914 (in CIJE).
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Educational Policy Analysis Archives; v7 n1-20 1999
EDRS PRICE	MF02/PC21 Plus Postage.
DESCRIPTORS	*Academic Achievement; Academic Standards; Charter Schools; Computer Uses in Education; Educational Administration; *Educational Policy; Educational Research; *Elementary Secondary Education; Foreign Countries; *Higher Education; Home Schooling; Internet; *Professional Development
IDENTIFIERS	South Korea

ABSTRACT

This document consists of downloaded copies of the 20 English-language and 3 Spanish language papers published in the electronic journal "Educational Policy Analysis Archives" during 1999: (1) "Ethnic Segregation in Arizona Charter Schools" (Casey D. Cobb and Gene V. Glass); (2) "Educational Research in Latin America: A Response to Akkari and Perez" (Mariano Narodowski); (3) "Effects on Students of a 4 X 4 Junior High School Block Scheduling Program" (R. Brian Cobb, Stacey Abate, and Dennis Baker); (4) "Less Truth than Error? An Independent Study of the Massachusetts Teacher Tests" (Walt Haney, Clarke Fowler, Anne Wheelock, Damian Bebell, and Nicole Malec); (5) "Some Comments on the Ad Hoc Committee's Critique of the Massachusetts Teacher Tests" (Howard Wainer); (6) "College Students' Use of the Internet" (Anna C. McFadden); (7) "Restructuring the Schoolbook Provision System in Indonesia: Some Recent Initiatives" (Dedi Supriadi); (8) "Scholastic Achievement and Demographic Characteristics of Home School Students in 1998" (Lawrence M. Rudner); (9) "High School Staff Characteristics and Mathematics Test Results" (Mark Fetler); (10) "Investigacion en educacion en America Latina: una continuacion del debate" (Abdeljalil Akkari, Soledad Perez); (11) "First Year Implementation of the School to Work Opportunities Act Policy: An Effort at Backward Mapping" (Arthur M. Recesso); (12) "Hiring Policy in United States Spanish Departments: Considerations of Social Class, National Origin, and Ethnicity" (Jerry Hoeg, Eric Cohen, and Christine Fullen); (13) "Contextualizing Home Schooling Data: A Response to Rudner" (Kariane Mari Welner and Kevin G. Welner); (14) "Supporting Community-Oriented Educational Change: Case and Analysis" (Linda Mabry and Laura Ettinger); (15) "Eleccion Educacional Entre Escuela Publica y Escuela Privada en Galicia: Un analisis comparativo del gasto de los hogares en las escuelas publicas y privadas" (Maria Jesus Freire, Jose Venancio Salcines); (16) "Race and Policy" (Ernest R. House); (17) "Autonomia por decreto? Paradojas en la redefinicion del trabajo del profesorado" (Jose Contreras Domingo); (18) "Demonstrated Actions of Instructional Leaders: An Examination of Five California Superintendents" (George J. Petersen); (19) "A Statist Political Economy and High Demand for

Education in South Korea" (Ki Su Kim); and (20) "Testing on Computers: A Follow-up Study Comparing Performance on Computer and on Paper" (Michael Russell). (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 431 820

VOLUME 7 - 1999

ISSN 1068-2341

Education Policy Analysis Archives

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

editor:
Gene V. Glass
Arizona State University

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

G. Glass

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

TM030000

BEST COPY AVAILABLE

Volume 7 1999

EDUCATION POLICY ANALYSIS ARCHIVES

Article Pages

1 Cobb & Glass: Ethnic Segregation in Arizona Charter Schools . . . 15
2 Narodowski: Educational Research in Latin America: A Response to Akkary and Pérez . . . 12
3 Cobb, Abate & Baker: Effects of 4X4 Junior High School Block Scheduling Program . . . 26
4 Haney, Fowler, Wheelock et al.: Less Truth Than Error?: An Independent Study of the Massachusetts Teacher Tests . . . 79
5 Wainer: Some Comments on the Ad Hoc Committee's Critique of the Massachusetts Teacher Tests . . . 6
6 McFadden: College Students' Use of the Internet . . . 7
7 Supriadi: Restructuring the Schoolbook Provision System in Indonesia: Some Recent Initiatives . . . 11
8 Rudner: Scholastic Achievement and Demographics Characteristics of Home School Students: 1998 . . . 39
9 Fetler: High School Staff Characteristics and Mathematics Test Results . . . 25
10 Akkari and Pérez: Investigación en educación en América Latina . . . 10
11 Recesso: First Year Implementation of the School to Work Opportunities Act Policy--An Effort at Backward Mapping . . . 49
12 Hoeg, Cohen & Fullen: Hiring Policy in United States Spanish Departments Considerations of Social Class, National Origin, and Ethnicity . . . 15
13 Welner & Welner: Contextualizing Homeschooling Data: A Response to Rudner . . . 15
14 Mabry & Ettinger: Supporting Community-Oriented Educational Change: Case and Analysis . . . 22
15 Freire & Salcines: Elección Educacional Entre Escuela Pública y Escuela Privada en Galicia . . . 26
16 House: Race and Policy . . . 17
17 Contreras: Paradojas en la Redefinición del Trabajo del Profesorado . . . 34
18 Petersen: Demonstrated Actions of Instructional Leaders: An Examination of Five California Superintendents . . . 28
19 Kim: A Statist Political Economy and High Demand for Education in South Korea . . . 29
20 Russell: Testing On Computers: A follow-up Study Comparing Performance On Computer and On Paper . . . 53

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 2872 times since January 14, 1999

Education Policy Analysis Archives

Volume 7 Number 1 January 14, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ
85287-0211

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Ethnic Segregation in Arizona Charter Schools

Casey D. Cobb
University of New Hampshire

Gene V Glass
Arizona State University

[The editorial review and decisions on this article were the responsibility of Anthony G. Rud Jr. of the Editorial Board.]

Abstract

Among the criticisms of charter schools is their potential to further stratify schools along ethnic and class lines. This study addressed whether Arizona charter schools are more ethnically segregated than traditional public schools. In 1996-97, Arizona had nearly one in four of all charter schools in the United States. The analysis involved a series of comparisons between the ethnic compositions of adjacent charter and public schools in Arizona's most populated region and its rural towns. This methodology differed from the approach of many evaluations of charter schools and ethnic stratification in that it incorporated the use of geographic maps to compare schools' ethnic make-ups. The ethnic compositions of 55 urban and 57 rural charter

schools were inspected relative to their traditional public school neighbors.

Nearly half of the charter schools exhibited evidence of substantial ethnic separation. Arizona charter schools not only contained a greater proportion of White students, but when comparable nearby traditional public schools were used for comparison, the charters were typically 20 percentage points higher in White enrollment than the other publics. Moreover, the charter schools that had a majority of ethnic minority students enrolled in them tended to be either vocational secondary schools that do not lead to college or "schools of last resort" for students being expelled from the traditional public schools. The degree of ethnic separation in Arizona schools is large enough and consistent enough to warrant concern among education policymakers.

Introduction

School choice arguably has become the most significant education policy issue of this decade. Choice programs such as vouchers, charter schools, open enrollment, and tuition tax credits continue to be discussed and debated at all levels of government and society. Charter schools are clearly at the forefront of the school choice movement, enjoying widespread public and legislative approval. Indeed, as of June 1998, 32 states have enacted legislation permitting the establishment of publicly funded charter schools.

Among the criticisms of school choice programs, and hence, charter schools, is their potential to further stratify schools along racial, socioeconomic, and other class-based lines (see e.g., Corwin & Flaherty, 1995; Elmore, 1987; O'Neil, 1996; Wells, 1993; Wells & Crain, 1992; Willms, 1986;). For instance, numerous commentators have expressed concern that charter schools will "skim" predominantly White, privileged students from public schools (see e.g., Buechler, 1996; Elmore, 1986; Fitzgerald, Harris, Huidekoper & Mani, 1998; Lee & Croninger, 1994; Wells, 1993). Were this to be true, charter schools could be found culpable of contributing to the re-segregation of America's schools. It is similarly plausible that charter schools could "cream" students of color, resulting in ethnically concentrated schools of choice. Given the novelty of charter schools and obstacles to obtaining relevant data, few empirical analyses have addressed these matters.

Proponents of charter schools consistently report that charters serve a proportionate (or sometimes higher) percentage of minority students in comparison to traditional public schools. Opponents say these data fly in the face of common sense-- that parents will tend to choose schools that predominantly serve children from backgrounds and class orientations similar to their own.

This study addresses two major questions within the context of ethnic stratification. First, is there evidence that charter schools are "skimming" White students? And second, are Arizona charter schools more ethnically concentrated than traditional public schools? The answers to these questions will help determine more generally if Arizona charter legislation (A.R.S. § 15-181) has resulted in increased ethnic segregation among its publicly funded schools.

Related Literature

Charter Schools and Ethnic Stratification

Several major charter school evaluations and policy reports concluded that the ethnic compositions of charter schools are in line with those of traditional public schools. Buechler (1996) reviewed various surveys, newspaper and magazine articles, research reports, and policy briefs from across the nation in compiling the 1996 report *Charter Schools: Legislation and Results after Four Years*. He summarized:

> As a group, the schools serve a student population comparable to the overall public school population in terms of race and socioeconomic status--not an elite population of upper-middle-class white students, as some had feared. Indeed, many charter schools have been designed explicitly to serve at-risk students. .... If anything, charter schools serve a more underprivileged student population than regular public schools do. (Buechler, 1996, pp. 26-27)

A Study of Charter Schools: First-Year Report, a comprehensive national evaluation sponsored by the U.S. Department of Education Office of Educational Research and Improvement, reported similar findings: "Charter schools have, in most states, a racial composition similar to statewide averages or have a higher proportion of students of color" (U.S. Department of Education, 1997, p. 24). This conclusion was based on state-by-state enrollment comparisons between a total of 214 charter schools and 21,656 public schools in ten states. Data were collected from the 1993-94 National Center for Education Statistics Common Core of Data. Charter schools in Michigan, Minnesota and Massachusetts served a higher percentage of predominantly (i.e., greater than 80%) minority students than did public schools. In contrast, charters in Arizona, California and Colorado enrolled a higher percentage of White (i.e., greater than 80%) students than did public schools. Overall, 95 of the 214 (44.4%) charter schools in the sample served student populations that were at least 80% White, while 45 charters (21.0%) enrolled student populations that were at least 80% minority.

On behalf of the Colorado Department of Education, the Clayton Foundation evaluated 24 Colorado charter schools in 1997 (Fitzgerald et al., 1998). Evaluators compared the percentages of students of color enrolled in charter schools with those of their sponsoring districts. Five charter schools out of the 24 served roughly (plus or minus two percentage points) the same percentage of students of color as their sponsoring districts. Four charters served a greater percentage of students of color than their sponsoring districts. In only one instance did the percentage of students of color (0.0%) served by the charter fall outside the range of percentages for district schools. The report concluded that, overall, charters enrolled racially diverse student populations.

Southwest Regional Laboratory published *Freedom and Innovation in California's Charter Schools* in the Fall of 1995. Surveying 54 of the 66 operating charter schools in California, evaluators asked administrators to estimate the percentages of racial and ethnic minorities that their schools served. To establish a comparison group, administrators were also asked to name nearby public schools that their students would have most likely attended had they not attended their charter school. Of the 83 public comparison schools identified by charter school administrators, 46 returned surveys that contained information on student characteristics. A comparison of the enrollments between charter schools and public schools led the evaluators to conclude that "the data do not support the hypothesis that charter schools are less racially balanced than nearby comparison schools" (Corwin & Flaherty, 1995, p. 112). Almost half of both the samples exhibited student populations comprising 50% or greater minorities. Further, only one in five charter schools served less than 20% minorities, an amount consistent with the comparison group.

An evaluation of Texas charter schools provided evidence of ethnic clustering (Taebel et al., 1997). Of the 17 charter schools in the study, nine were 90% or higher ethnic minority. Eight of these had curricula designed to serve at-risk students. The non-at-risk schools enrolled over three-fourths of all White students served by the charter cohort.

Exclusionary Admissions Practices

There is some concern that selective admissions policies could contribute to racial imbalances among schools. The Colorado Department of Education (Fitzgerald et al., 1998) found no evidence, at least "on the surface," of exclusionary practices. All, save for one, of the 24 charter schools in the study used some random process, such as a lottery, or a first-come-first-served policy to admit students. The lone exception was the Stargate Charter School, which targeted gifted and talented students. For students qualifying as intellectually or academically gifted, the school allocated the first 100 seats--with reserved race and gender slots based on district percentages--on a first-come-first-served basis. The remaining 50 seats were allocated by lottery. Interestingly, Stargate enrolled 12% students of color in a district with schools that ranged from a low of 12% to a high of 60% minority enrollment.

Fieldworkers for the U.S. Department of Education (1997) conducted several telephone surveys, site interviews, and focus groups with charter school directors. Of those surveyed by phone, nearly three-quarters indicated that applications for admission exceeded capacity. For those schools with waiting lists, 39% reported using some random selection process, 41% employed a first-come-first-served policy, 10% used some combination of these policies, and the remaining 10% used some "other" [emphasis in the original] process. Although the evaluators did not find evidence of explicit discriminatory admissions practices, they remarked in an endnote:

> More subtle processes of selecting students, however, may be at work. Intensive field research in subsequent years should allow us to probe deeper into selection processes. For example, we will want to ask, in situations where it is possible, whether charter schools actively seek out students from diverse ethnic or racial backgrounds. The research team documented several cases where the schools do reach out actively, but we cannot report definitive data at this time. (U.S. Department of Education, 1997, p. 47)

Nine of seventeen Texas charter schools exhibited acute cases of racial distinctiveness (Taebel et al., 1997). Evaluators attributed the enrollment imbalance to four factors, two of which were a first-come-first-served admissions policy and word-of-mouth marketing. Indeed, parents cited word-of- mouth as the most influential form of advertising. The evaluators commented:

> While it is reassuring to know that parents share such information with one another, there is a danger of exclusion when recruitment is a function of whom you know. "Friend or relative" communication networks also tend to be homogeneous with respect to race and class. Relying solely on this kind of communications for student recruitment means that those who come first may be racially and socioeconomically similar to the existing student body. (p. 97)

It is not uncommon for charter schools to require parents to sign formal "involvement" agreements to participate in their child's instructional programs. Such contracts have the potential to serve as sorting mechanisms, excluding parents who may be willing but are practically unable to fulfill such commitments. Corwin and Flaherty (1995) noted poignantly, "Although charter schools were created to allow parents greater choice in the kinds of schools their children attend, parent contracts seem to give schools greater choice over the kinds of parents they choose to serve" (p. 105).

Becker, Nakagawa, and Corwin (1995) asked 28 charter school administrators in California what factors they considered in accepting new students. Twenty-five percent indicated that an "essential" determinant was that the "parent or guardian will participate in requested ways" (p. 18). From their original sample of 34 charter schools, 27 (79%) reported using parent involvement contracts.

Methodological Issues

Many of the national policy reports and evaluations lack the sophistication and rigor necessary to draw valid conclusions about the possible segregating effect of charter schools. In the first, there is great risk in making sweeping statements about charter schools given the variability in state charter school laws. Some states carefully regulate the admissions process while others do not. States also differ widely in terms of the restrictions on the number and types of charters to be awarded. For example, legislation in over a third of the charter states either encourages or requires a portion of charter schools to appeal to the needs of at-risk youth (Buechler, 1996).

Second, data aggregated at the state and even district level mask variation among schools. For instance, the U.S. Department of Education (1997) reported that in 1995-96, Arizona charter schools served 20.2% Hispanic students while the public schools served 27.6%. These aggregated data cannot speak to the variability in the percentage of Hispanics served within either segment. Several charter and public schools in Arizona are ethnically concentrated, but this information is shrouded in grossly aggregated statistics.

Finally, difficulty in obtaining accurate data is a common complaint among charter school researchers. This is not altogether surprising, as by design one of the major advantages of charter schools is to free them from burdensome record keeping responsibilities. For example, Corwin and Flaherty (1995) asked traditional public and charter school administrators to estimate the percent of minorities that their school enrolled within very broad ranges (i.e., between 0-19%, 20-49%). Obviously, imperfect data attenuate the strength of evaluators' conclusions.

School Choice and Social Stratification

Given the dearth of empirical studies that address charter schools and ethnic stratification, the literature review was broadened to include studies on school choice and social stratification. Considerably more research has been conducted in this area.

Since the United Kingdom passed public school choice legislation in 1980, it has served as the focus for many studies on parental choice (Willms, 1996). Adler, Petch, and Tweedie (1989) asked over 600 parents in Scotland to identify their criteria for choosing a school. They found that few parents emphasized educational considerations, such as curriculum or test results. Instead, their main reasons for choosing a school were based on social factors, such as school climate and general reputation, as well as with practical issues, such as proximity.

Witte (1993) interviewed 171 parents who participated in the Milwaukee Choice Program in 1991. Although the most emphasized criteria for selecting a school was perceived educational quality, 75% of the parents considered the "other children in chosen school" to be an important or very important factor in their decision. Incidentally, 80% deemed location of chosen school important or very important.

Based on a nationally-representative sample of secondary students in the U.K., Echols, McPherson, and Willms (1990) reported that choice schools tended to serve populations of above average socioeconomic class. In addition, those parents who exercised choice were relatively more educated and belonged to a higher social class. Willms (1996) conducted a more sophisticated longitudinal analysis to investigate the extent to which Scottish communities had become socially segregated. He reported that "there was clearly greater propensity to exercise choice among higher social class and better educated parents" (p. 142) and that "parents choosing within the state sector disproportionately chose schools with higher mean SES than other state-sector schools" (p. 143).

Based on a nationally representative sample of secondary students in the U.K., Echols, McPherson, and Willms (1990) reported that choice schools tended to serve populations of above average socioeconomic class. In addition, those parents who exercised choice were relatively more educated and belonged to a higher social class. Willms (1996) conducted a more sophisticated longitudinal analysis to investigate the extent to which Scottish communities had become socially segregated. He reported that "there was clearly greater propensity to exercise choice among higher social class and better educated parents" (p. 142) and that "parents choosing within the state sector disproportionately chose schools with higher mean SES than other state-sector schools" (p. 143). Whitty (1997) conducted an extensive review of school choice research in England, New Zealand, and the United States. Within the English system, Whitty observed that parental choice did not lead to a "truly diversified system" (p. 14) and Walford (1992) concluded that choice will "discriminate in particular against working class children and children of Afro-Caribbean descent" (p. 137). A major study on school choice in New Zealand reported similar polarizing effects. Whitty (1997) ultimately summarized, "...my conclusion from the evidence we have to date is that, far from being the best hope for the poor, as Moe (1994) suggests, the creation of quasi-markets is likely to exacerbate existing inequalities" (p. 5).

A two-year study on school choice programs in New York, Philadelphia, Boston and Chicago revealed that minority students and students from low-income families were underrepresented by choice schools with selective admissions policies (Moore & Davenport, 1990). The authors reported:

> In these school systems, school choice has, by and large, become a new improved method of student sorting, in which schools pick and choose among students. In this sorting process, black and Hispanic students, low-income students, students with low achievement, students with absence and behavior problems, handicapped students, and limited-English- proficient students have very limited opportunities to participate in popular-options high schools and programs. Rather, students at risk are disproportionately concentrated in schools where their fellow students are minority, low-income, and have a variety of learning

problems. (p. 188)

Methods

Data Sources

October enrollment data disaggregated by race and ethnicity, gender, and grade level for the years 1994-1997 were obtained from the Arizona Department of Education (ADE) School Finance Division for all public elementary and secondary schools in Arizona. The same data were acquired from the ADE for charter schools for the years 1995-1997. Enrollment figures in these schools comprise the entire corpus of data.

All public schools, including charter schools, are required to report October 1 enrollments by race and ethnicity, gender, and grade level (John Eickman, personal communication, May 26, 1998). The racial and ethnic codes used by the ADE are White, Black, Hispanic, American Indian or Alaskan Native, and Asian or Pacific Islander.

The ADE collected October enrollment data from 51 charter school sites in 1995 (charter schools' inaugural year), 132 in 1996, and 137 in 1997. Although one would assume that the number of charter schools reporting enrollment data would represent the total number of operating charters for that year, this is not the case. Conversations with several members of the ADE failed to confirm precise numbers of operating charter schools. Charter schools open and close during the year, and do not necessarily open in the year that they are granted, thus making it difficult to maintain exact numbers of operating charters. Best estimates from ADE dated lists of charter schools are displayed in Table 1.

Table 1
Number of Charters Reporting Enrollment Data and Estimated
Number of Operating Charters

Year	No. Reporting Data	Est. No. Operating
1995	51	51
1996	132	135
1997	137	215

Most notable are the October enrollment data submitted by charter schools for 1997, which deviate substantially from the often reported 240 to 260 operating charter schools in the third year of their existence. (The 1997 October enrollment data were collected by the ADE as late as May 15, 1998, which allowed sufficient time for schools to report. The ADE Charter Schools Handbook mandates that all schools report these data by October 31 of each year.) To obtain the number of operating charter schools in the 1997-98 school year, a team of researchers (the author, Gregg Garn and Linda Brock-Nelson) queried the 250 charter schools listed by the ADE as of March 23, 1998. Results indicated that at most 215 charter schools were in operation in the Spring of 1998. Thus, 1997-98 charter school enrollment data used here represents roughly two-thirds of the population of operating charter schools.

Schools not classified as regular public schools (e.g., accommodation schools, vocational and technical schools not operated by public school districts, and the like) were removed from the analysis. For instance, 34 of these non-traditional schools that served 6,100 students were eliminated from the 1996 data set.

Digital map data of metropolitan Phoenix street grids, census tracts, and zip code boundaries were acquired from the data archives of the Arizona State University Geographic Information Systems (GIS) Laboratory. Site addresses of charter and traditional public schools were obtained from the ADE School Finance Division. In those instances where addresses were absent or in a form that did not indicate geographic location, the ADE School Report Card website (http://sais.ade.state.az.us/rcweb/) or direct inquiries to schools provided street addresses. In all, 586 addresses were geocoded onto a digital map using Arcview. Initially, about three-fourths of the addresses were successfully matched by Arcview. The remaining 136 addresses were manually plotted by reference to the 1998 edition of the *Phoenix Metropolitan Street Atlas*. Lastly, selected census data were acquired from the Maricopa Association of Governments (MAG). These data were collected by MAG as part of the 1995 Special Census of Maricopa County.

Procedures

ADE October enrollments for the years 1995-1997 were aggregated by year for all schools. First, ethnic distributions were compared between charter and traditional public schools. Then, for reasons explained later, comparisons were made after removing reservation charter schools.

The remainder (to be sure, the core) of the analysis involved a series of comparisons between the ethnic compositions of adjacent charter and public schools in Arizona's most populated region and its rural towns. This methodology differs from the approach of many evaluations of charter schools and racial stratification in two major respects. First, it examines the potential for ethnic separation at the disaggregated level of school. Second, for half of the charter schools, explicit comparisons are made within the context of geographic maps.

To see the ethnic separation in Arizona charter schools, one must examine the geography of the situation. The crucial question is not what percents of ethnic groups either are or are not in charter schools; rather, the crucial question is how are ethnic groups distributed between propinquitous charter and traditional public schools. This question is addressed differently in small rural places and in large metropolitan areas. In the former, because attendance catchment areas are small, it is sufficient merely to list small towns that have charter schools and compare their ethnic composition to the traditional public school or schools in the same town. In the case of large metropolitan areas, it is necessary to plot actual maps of these areas and inspect the ethnic distributions of adjacent charter and traditional public schools.

Attempts to depict the magnitude of differences among schools' ethnic compositions while holding constant size and grade level *through various statistical measures* prove problematic. Popular measures of level of segregation, such as the Dissimilarity Index, and measures of equity, such as the Gini coefficient or Lorenz Curve, are highly sensitive to numbers of students in schools. The relative smallness of charter schools makes comparisons via these types of measures questionable. Moreover, within this context, these indices are simply powerless to detect between-school segregation. No statistical technique can aptly discern differences among urban schools as completely as maps.

These analyses are exploratory (Tukey, 1977), not confirmatory. It is impossible in advance of studying these data in detail to specify individual "hypotheses" to test. Hence the exploratory nature of these analyses. There are no significance tests here simply because there is no sampling of a probabilistic sort that could give meaning to any probabilistic inferences. Absent also are correlational techniques such as multiple regression analysis, which decontextualize the data and do not provide adequate means to detect the existence of a phenomenon, if it exists; further, they are difficult for laypersons to interpret.

Map Analysis of Urban Charter Schools

Using Arcview, pertinent Arcview coverage and shape files, and ADE school address data, charter and traditional public schools were plotted onto a digital map of metropolitan Phoenix. Each school was virtually linked to 1996 October enrollment data. The 1997 data were not plotted because they were not available at the time, and even so, were not nearly as complete as the 1996 data.

Data were available for 55 charter and 518 traditional public schools in metropolitan Phoenix. Metropolitan Phoenix accounts for nearly 60% of Arizona's population. In the Fall of 1996, metropolitan Phoenix was home to about half (47%) of the 132 charter schools in the state.

Arizona is demographically unique in that two urban centers account for the majority of its populace. Metropolitan Phoenix and the city of Tucson comprise over three-fourths of the state's population. In the interest of time, and considering that metropolitan Phoenix is over three times as populated as Tucson and is home to far more charter schools, Tucson was excluded from the analysis.

The exploratory nature of the map analysis ultimately led to a systematic approach with which to search for ethnic separation. The ethnic composition of every charter school in metropolitan Phoenix was compared to that of nearby traditional public schools of comparable grade levels. This spatial analysis was done using maps that ranged in coverage from 5 to 28 square miles. In some areas, multiple charter and multiple public schools coexisted. In others, a single charter school was located in the vicinity of five or six public schools. Judgments were made as to the presence and degree of ethnic separation primarily on the basis of the magnitude of difference in the proportion of White students enrolled. Typically, occurrences of ethnic separation were documented in instances where the magnitude of difference was 15% or greater. Multiple schools of various sizes, grade levels, distances apart, and ethnic distributions complicated matters, but were factors all of which were carefully considered. The nature of "nearby" is what remains to be unpacked and will surely be contested by those who advance other explanations of the findings discovered here. Judging whether a traditional public school is "nearby" a charter school and hence may serve as a comparison of enrollment data is a complex judgment not captured simply by geographic distance (i.e., miles separation), school district boundaries or other obvious and easily specified criteria. For example, canals, cultural factors like the fact that Mesa is Mormon in many areas, sections of cities isolated by freeways or mountains, and differences in population densities must be simultaneously considered when making these judgments.

For the most part, the analysis relied on the maps prima facie. But there is doubtless a story behind each picture that could not be told here. Given the large number of charters addressed by this study, it seemed unreasonable to try to account for all potential alternative hypotheses (that is, alternative to attributing ethnic separation to a charter school). In ambiguous instances or otherwise where deemed useful, additional information was provided to supplement the face value information provided by the maps. Certainly, the core of the analysis was spatially and numerically based, but where applicable, ancillary evidence provided further explanation.

The map analysis spawned a less comprehensive but more clearly specifiable and readily interpretable matched comparison analysis. The nearest public school or schools, again of comparable grade level, were matched with each charter school. The geographically related comparison group was chosen in preference to a random sample of public schools in Maricopa County because it controls, in effect, for geographic location and ethnic composition of the immediate region. Indeed, a random sample would not be prudent because charter schools do not locate under the same conditions that traditional public schools do. In those cases where the grade range of a charter school was not mirrored by a nearby public school, public schools that "covered" the grade levels were combined. For example, for a K-12 charter school, the nearest K-5, 6-8, and 9-12 public schools collectively served as the matched comparison. In all, the ethnic distributions of 55 matched pairs were inspected. In addition, charter and public comparison schools were grouped into three categories: schools below 30% White, over 70% White, and in between 30-70% White.

Lastly, mere surface level exploration of the data raised suspicion of a relationship between the educational mission of charters and their ethnic make-up. This triggered the categorization of secondary level charters into either college prep or voc-ed programs. Classifications were primarily based on self-described school missions, organizations and philosophies, and instructional programs found in the 1996 online

ADE School Report Cards (http://sais.ade.state.az.us/rcweb/). Descriptors such as "at-risk," "school-to-work," and "tech-prep" placed schools in the voc-ed category. Indicators of a more mainstream or college-bound program (e.g., "academic college preparatory," "college prep," or "accelerated learning") designated schools as college prep.

Analysis of Small Town Charter Schools

The rural data are inclusive of small towns that contain public schools and at least one charter school. This straightforward analysis compared charter schools to traditional public schools of the same grade level. Additionally, the analysis that explored the relationship between educational program and ethnic composition among urban charters was repeated for the rural cohort. A total of 57 rural charter and 88 public schools (which included several reservation schools) from 36 rural Arizona towns was examined.

In sum, the ethnic compositions of 55 urban and 57 rural charter schools were inspected relative to their traditional public school neighbors.

Results

Tables 2-4 present aggregated ethnic distributions of charter and traditional public schools for the years 1995-1997. Across all years, charter schools enrolled a considerably higher proportion of Black students than traditional public schools. In contrast, Hispanic students were significantly underrepresented in charter schools. For instance, in 1996, Hispanic students participated in charter schools at half the rate at which they participated in traditional public schools. That same year witnessed a three-fold increase in American Indian charter school participation over their presence in traditional public schools. This is commented on below. Also notable are the percentages of White students served by charter and traditional public schools, which differed only marginally for the first two years. By the third year of their operation, however, charter schools enrolled a higher percentage of White students than the traditional public schools. (An important caveat: it should be noted once again that for reasons unknown the charter school enrollment data for 1997 were much less complete than for the prior years.)

Table 2
1995 Ethnic Compositions of Arizona Schools

	Trad. Public Students (n=1159)		Charter Students (n=51)	
Ethnicity	**No.**	**%**	**No.**	**%**
White	434,473	57.6%	3,718	59.9%
Black	31,132	4.1%	631	10.2%
Hispanic	222,751	29.5%	1,215	19.6%
Am. Indian	52,868	7.0%	564	9.1%
Asian	12,957	1.7%	79	1.3%
All	754,181	100.0%	6,207	100.0%

Note. Due to rounding, percents do not sum exactly to 100.0%

Table 3
1996 Ethnic Compositions of Arizona Schools

	Trad. Public Students (n=1148)		Charter Students (n=132)	
Ethnicity	**No.**	**%**	**No.**	**%**
White	440,894	56.8%	9,776	55.2%
Black	32,264	4.2%	1,251	7.1%
Hispanic	236,475	30.4%	2,919	16.5%
Am. Indian	53,527	6.9%	3,567	20.1%
Asian	13,712	1.8%	213	1.2%
All	776,872	100.0%	17,726	100.0%

Note. Due to rounding, percents do not sum exactly to 100.0%

Table 4
1997 Ethnic Compositions of Arizona Schools

	Trad. Public Students (n=1181)		Charter Students (n=135)	
Ethnicity	**No.**	**%**	**No.**	**%**
White	440,887	55.9%	11,804	61.4%
Black	33,521	4.3%	1,176	6.1%
Hispanic	245,528	31.1%	3,442	17.9%
Am. Indian	53,905	6.8%	2,484	12.9%
Asian	14,461	1.8%	307	1.6%
All	788,302	100.0%	19,213	100.0%

Note. Due to rounding, percents do not sum exactly to 100.0%

Some questions have been raised about the nature of charter schools on American Indian reservations. Nearly all of them are converted from erstwhile reservation schools that were formerly funded by BIA or other federal programs. Given their geographic isolation and virtually unchanged condition, reservation charter schools do not offer genuine educational alternatives for students in those communities.

If the reservation schools are removed from Tables 2-4 (i.e., if all schools for which the American Indian percent of students is 93% or greater are taken out) the data are even more revealing of the segregation trend emerging in the charter schools. For 1996, the difference in the percentage of White students attending charter and traditional public schools widened nearly ten-fold after removing the reservation schools (see Tables 3 and 6). For 1997, the difference nearly doubled (see Tables 4 and 7).

Table 5

1995 Ethnic Compositions of Arizona Schools with Reservation Schools Removed

	Trad. Public Students (n=1105)		Charter Students (n=50)	
Ethnicity	**No.**	**%**	**No.**	**%**
White	434,074	59.6%	3,718	64.8%
Black	31,117	4.3%	631	11.0%
Hispanic	222,675	30.6%	1,215	21.2%
Am. Indian	27,543	3.8%	91	1.6%
Asian	12,928	1.8%	79	1.4%
All	728,337	100.0%	5,734	100.0%

Note. Due to rounding, percents do not sum exactly to 100.0%

Table 6

1996 Ethnic Compositions of Arizona Schools with Reservation Schools Removed

	Trad. Public Students (n=1092)		Charter Students (n=124)	
Ethnicity	**No.**	**%**	**No.**	**%**
White	440,519	58.7%	9,760	67.0%
Black	32,250	4.3%	1,248	8.6%
Hispanic	236,409	31.5%	2,916	20.0%
Am. Indian	27,202	3.6%	446	3.1%
Asian	13,683	1.8%	208	1.4%
All	750,063	100.0%	14,578	100.0%

Note. Due to rounding, percents do not sum exactly to 100.0%

Table 7

1997 Ethnic Compositions of Arizona Schools with Reservation Schools Removed

	Trad. Public Students (n=1124)		Charter Students (n=129)	
Ethnicity	**No.**	**%**	**No.**	**%**
White	440,521	57.8%	11,792	68.4%
Black	33,511	4.4%	1,176	6.8%
Hispanic	245,453	32.2%	3,440	19.9%
Am. Indian	28,068	3.7%	533	3.1%
Asian	14,441	1.9%	304	1.8%
All	761,994	100.0%	17,245	100.0%

Note. Due to rounding, percents do not sum exactly to 100.0%

Finally, not only are the charter schools disproportionately White, the trend to become even more White can be seen by inspecting all three years' data (see Table 8). (Once again, the 1997 data should be interpreted with caution as these are curiously incomplete.)

Table 8
Percent White Students in Charters (Excluding Reservation Schools)

Year	% White
1995	64.8%
1996	67.0%
1997	68.4%

Aggregated data, like those presented above, are powerless to illuminate potential ethnic separation at the level of school. For instance, in 1996, well over half (56.3%) of the Black students attending charter schools were served by just three schools. As the data are explored even more (here, and in subsequent analyses), the trend toward ethnic stratification becomes clearer. The map and small town analyses provide the best opportunity for discovering ethnic separation in urban and rural communities if it exists.

Maps of Urban Charter Schools

Nineteen maps (Figures 1-19) of sections of metropolitan Phoenix contain 34 different charter and 128 different traditional public schools. The maps averaged a charter-to-traditional school ratio of 1:5.2. Together, they covered 220 non-duplicated square miles in six cities.

The maps are rich with information, conveying spatial relationships among schools and unique geographic properties such as canals, rivers, and major streets and highways. They include the

• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 573 times since January 23, 1999

Education Policy Analysis Archives

Volume 7 Number 2 January 23, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ
85287-0211

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Educational Research in Latin America: A Response to Akkary and Pérez

Mariano Narodowski
Universidad Nacional de Quilmes

Abstract In a recent article which appeared in *Educational Policy Analysis Archives* 6 (7) 1998, "Educational Research in Latin America: Review and Perspectives," Abdeljalil Akkari and Soledad Pérez envisage carrying out a general analysis of the situation of educational research in Latin America, in an attempt to describe the context of its historical formation. They focus on the main theoretical framework, they identify the principal institutions involved in educational research, and consider the priorities for future research in the region.

This article does not claim to provide a global reply to the work presented, although it does hope to elaborate upon certain aspects, to clear up matters which are presented in an over-simplified manner, and to extend the analysis in order to capture the intricacies of the problem in all its complexity. The *Education Policy Analysis Archives*, considered one of the top level journals, is consulted by a wide range of international readers who often possess only superficial knowledge of educational research in Latin America

Defining "Latin America"

The work of Abdeljalil Akkari and Soledad Pérez attempts to analyze the development of educational research in Latin America. In the article, the authors seem to take for granted a sort of homogeneity among Latin American countries. The conclusions they reach never specify differences or contrasts which exist among these countries: the article creates a Latin American panorama without bothering with factors specific to each country, and even takes for granted the existence of some sort of "identity" as regards educational research.

In our opinion, this analysis deserves a more rigorous approach. That which is usually termed "Latin America" includes more than twenty independent nations which, despite sharing many common characteristics, display many varying features. Among these are: different languages (Spanish, Portuguese, as well as another several dozen existing languages of the native indians), differences in terms of systems of government (for example, presidential systems in Argentina and Mexico; a semi-presidential system in Brazil; "socialist democracy" in Cuba, etc.); ethnic, immigrant, political, economic, cultural, and even territorial differences. One must bear in mind the war-like confrontations between Argentina and Chile; Colombia and Venezuela; or the latest one between Peru and Ecuador, only to mention those which are most recent. Within the limits of this article, it would certainly be impossible to cover each and every one of the similarities and differences which exist among Latin American nations.

From the educational, academic, or scientific point of view, there are still more differences, and the supposition of an identity of processes calls for closer examination. There are countries with universities which have existed for more than four hundred years, while others have only recently established universities. There are countries where some investigators have been awarded Nobel Prizes in science. There are countries where the growth in the indices of basic schooling are similar to those of the most highly developed European countries, while in some others, the indices are the lowest in the world.

One could take as an example the Latin American countries from Central America, where the situation would appear to be homogeneous. Nevertheless, certain studies have shown the differences which exist in university development between Costa Rica and the rest of the Central American countries in terms of the different dimensions of academic activity (Levenberg, 1995). Disparities have also been found among Caribbean nations, some of which are not "latin" (Dachary and Burne, 1995).

As regards science and higher education, the situations are very heterogeneous. According to data from the World Bank, between 1992 and 1995, in comparison with Korea, Argentina produced more than twice the number of publications in international scientific journals. As for Mexico, during the same period the amount it produced was similar to that of Argentina, but both of them produced half the number of papers from Brazil and only 1% of those from the United States. One further piece of data substantiates the diversity of the experiences: Latin American countries such as Paraguay, Guatemala, and Honduras can barely sustain research programs, and their presence in the international context is minimal (Almada, 1979). In the words of Garcia Guadilla (1998: pp. 432-433), and in reference to research investigation on higher education:

> "Although a small group of countries has achieved initial institutionalization, the rest of the countries still lack the basics needed to develop this area as a field of study, due to a limited production of research as well as to the nonexistence of research centers and/or specific training to develop autonomously in this area of knowledge

These examples are not directed at refuting the possibility of the existence of a single identity as regards Latin American educational research, but to advance in the idea that that single identity should be more the result of the evidence of an investigation and less an assumption providing a starting point for further work. In fact, in other works we have emphasized the need for the development of comparative studies in Latin American educational research which allow for the recognition of both the national peculiarities and the regional identities (Martinez Boom and Narodowski, 1998).

In the Akkari and Pérez study, the lack of a more thorough investigation of these peculiarities results in the analysis of common processes. The case of Mexico is a perfect example of this as in the article it is mentioned only once and there, to name a non-governmental organization dedicated to educational research. The article neglects the work carried out in prestigious and traditional institutions dedicated directly or indirectly to educational research such as the Universidad Nacional Autonoma de Mexico, the Colegio de Mexico, or the Departamento de Investigaciones Educativa of the Instituto Politécnico Nacional . This omission is not irrelevant considering that Mexico is the Latin American country with the greatest proportion of investigators carrying out their activities in public institutions.

This lack of specification and its hurried generalization produce results which are equally partial and, to say the least, questionable. The analysis of the situation of educational research in Latin America calls for the development of local studies and comparative examination which has barely begun. Obviously, this response hopes to begin to complete the work of Akkari and Pérez and not to substitute it but rather to point out, schematically, certain areas which require deeper examination.

History and Theoretical Structures

There are no studies which analyze the history of educational research in Latin America in a global and detailed manner. Despite this, we can point out that diverse experiments in educational research began around the end of the nineteenth century and the beginning of the twentieth and did not merely begin in the decade of the thirties. Furthermore, Argentina was able to achieve a level of excellence in its investigators in education when, in 1914, the University of La Plata created the Laboratory of Experimental Pedagogy (Dabat, 1992). In Brazil, the first study of the history of education ,of Pires de Almeyda, was published in 1889. Therefore, sufficient evidence does not appear to exist which would affirm that in the region, educational research "...most research has primarily been descriptive rather than empirical or applied." In fact, in Latin American academic institutions, a tradition still exists which recognizes research only if it is of an empirical or even experimental nature. In Mexico, the implementation of courses for the masters and doctors degrees in the University of Mexico in 1955, indicated a clear emphasis upon the training of investigators in education with specific stimulus towards empirical projects. (Ducoing, 1990).

The reactions provoked by this tradition would appear to be proof of the importance of applied and experimental research in Latin America. Towards the end of the 1920's, the educational philosopher, Juan Manovani, in a journalistic interview, complained of the excess of experimentalism in educational research, criticizing psychology and experimental didactics and what he termed "the reign of the method". (Carli, 1955).

On the other hand, in several countries, Colombia, Mexico, Brazil, and Argentina, starting in the 1930's, the development of the educational system generated great challenges in educational research, especially the trend towards didactic methods and active or radicalized positions in psychology which resulted in a high number of Latin American participants in the Ligue International de l'Ecole Nouvelle and their active participation in international discussions which took place there. This urgency to improve educational practices generated a tendency towards research applied to teaching methods, as in the case of the Instituto de Investigaciones Educacionales (in Brazil), under the direction of Anixio Texeira, a pedagogue who was able to combine political management with research and teaching (Gonçalves Vidal, 1996).

This rich and heterogeneous history prevents a quick categorization of the principal theoretical structures or of those which have played a more important role in educational research. It is true that the three theoretical tendencies pointed out in the paper of Akkari and Pérez were and are notably relevant: the one created by the ECLAC, the so-called Theory of Dependence, and the pedagogy of Paulo Freire. However, several relevant theoretical positions have maintained and/or currently maintain a central position in the Latin American academic debate and must not be discarded.

Establishing significant theoretical tendencies implies in the first place carrying out the detailed job of collecting and checking all the Latin American publications related to educational research, a task which has barely begun and presents many difficulties. For example, a study was recently published which provided a list of existing studies on the Argentine educational system from within the period 1976-1994 (Feldfeberg, 1995). This work, of unquestionable bibliographical value, lists 284 published studies. Moreover, it leaves out the educational research of the rest of the countries and even Argentine educational research itself dedicated to other subjects, besides the obvious (and perhaps unavoidable) omissions of the study. A quick perusal of the listed studies shows that a categorization of all these works greatly exceeds the theoretical inclinations suggested by Akkari and Pérez, and one could even affirm that none of these 284 could be placed within the tradition founded by Paulo Freire, one of the theories considered of great importance by Akkari and Pérez.

Something similar occurs with the annual meetings of the Associação Nacional de Pesquissa e Posgraduacao em Educaçao (ANPED and not ANPE as it appears in the Akkari and Pérez article). There alone, hundreds and hundreds of papers are presented by the participants. To determine the principal theoretical trends, the shades, and the ideological slant of the investigations calls for research which has only just been begun by Aedo-Richmond (1996)

For this reason, great features of theoretical structures of importance in Latin American educational research will now be presented which are or have been part of its development. Due to the circumstances described, the aim is not to deal with the totality of the tendencies but only to contribute to the completion, although if only partially, of an extensive, heterogeneous panorama.

In the field of sociology of education, the functionalist and structural- functionalist positions have generated numerous studies which provide empirical evidence about different aspects which are relative to the schooling process. Some of these studies have been influenced by the works of CEPAL while others have maintained an orthodox point of view (Tedesco, 1987). The Marxist and reproductivist points of view, influenced as much by Latin American Marxism as by the Althusserian critical trends, reached their peak between the 1970's and the beginning of the 1980's (Labarca and others, 1977; Vasconi, 1988).

Theoretical and methodological ruptures have been occurring in this field for more than three decades. As regards the theoretical side, we should point out the current popularity of multicultural studies and of those focused on categories such as gender, ethnicity, or class. It is impossible to name all the authors who adhere to this position. It is worthwhile mentioning, as an illustration, a recent book edited in Brazil comprising original work of European, North American, and Latin American authors working along this line of research (da Silva, 1998). The exclusion of this tendency is misleading considering that this tends to cause one to think that in the 1990's in Latin America, post-structuralist discourse prevails in most of the educational research.

Also of great importance are the studies undertaken by several groups which base their research on the contributions of M. Foucault. In Colombia, an inter-university group developed which has produced books and papers from this perspective (Diaz Villa, 1993). These contributions, and the theoretical tools of "critical pedagogy" have also been utilized in the curriculum field, as can be seen in numerous studies produced in several countries - especially in Mexico (Diaz, 1989).

From the methodological point of view, modern positions such as the ethnographic approach attempted by many investigations, particularly in Mexico and Chile are worth pointing out. Both the empirical studies as well as the epistemological proposals about the utilization of ethnology in educational research arising in this area are very well known. The so- called participatory research must also be mentioned (Rockwell, 1991).

A fundamental chapter in the present theories which have the most influence upon educational research is given by the constructivist theory, based on the work of Jean Piaget in the field of educational psychology and didactics. The investigations of E. Ferreiro and A. Teverosky have been the center of a huge debate and have been published in books and re-edited many times. They constitute one of the theoretical positions most akin to the daily reality of teachers. The Brazilian journal *Educação & Realidade* has devoted several issues to the debate around constructivism, its epistemological foundations, the evidence that it produces in research, and its consequences in educational practice. In addition, one should not discard the importance of many research groups dedicated to cognitive psychology applied to education.

Educational research in the field of the history of education has been one of the most productive types of investigation in the last decades. Especially in the 1990s, journals and scientific societies dedicated to this field of study have appeared, and four international congresses on Latin American Educational Historiography have taken place. Some authors already propose a positive balance of this experience (Sanchez Gamboa, 1996) although it would not appear easy to reach generalized theoretical conclusions in meetings where more than four hundred papers are presented. As is asserted by Tellez (1996, p. 10), one cannot aspire to a unity of points of view but at best to generate "a map which allows us to create a view both of the diversity of matters which occupy the investigators' attention...and of the keys of research which are involved."

In several Latin American countries, a very important field of research exists which has to do with the relationship between the State and education (Torres and Puiggrós, 1995). It is probable that works on the politics, administration, financing and economics of education are those where it is the easiest to detect theoretical influences and where these are the easiest to generalize. Studies generated by governmental, inter- governmental, and international organisms have given a definite slant to educational research to such a degree that in another study we have proposed the existence of a sort of "State pedagogue" in Latin American countries (Narodowski, 1997). However, it is not easy to determine the theoretical contents of the positions presented, especially at the time when divergences or peculiarities between apparently congruent theoretical positions are detected.

To sum up, this examination of Latin American educational research reveals a large amount of divergent studies and diverse theoretical and methodological positions, where a proliferation of metatheoretical studies can be observed. Furthermore, this review shows that the theoretical tendencies which dominate in the different stages of the development of educational research are the same as those which predominate in Europe or the United States.

If we set aside the analysis of the periods when the Latin American investigators had limited freedom due to the presence of military dictatorships (which in the last ten years have not played a role in almost any of the countries of the region) (Braslavsky, 1991), we can assert that the principal theoretical tendencies appear in the different geographical areas. This situation appears to have been accentuated in the last decade thanks to the popularization and reduction in cost of technology for the transmission of information such as Internet.

Obviously, it remains to be investigated whether the characteristics and use of these theories have a particular bias which constitutes some sort of single identity in all of Latin America or in some of its countries

References

Aedo-Richmond (1996), Ruth. Education in Latin America: A Selected Bibliography (1986-1995). *Compare*; v26 n2.

Almada, Martín (1979) *Paraguay: educación y dependencia*, Buenos Aires, 1974.

Akkari, Abdeljalil & Pérez, Soledad. (1998). Educational Research in Latin America. Review and Perspectives. *Educational Policy Analysis Archives*, Vol. 6, No. 7.

Braslavsky, Cecilia (1991). Education at a Time of Democratic Transition in South America. Prospects; v21 n4..

Carli, Sandra (1995) *Entre Ríos escenario educativo 1883- 1930*, Cuadernos, Paraná.

Da Silva, Luiz (ed.) (1998) *A escola cidada no contexto da globalizaçao*, Vozes, Rio de Janeiro.

Dabat, Roque (1992) *El positivismo argentino. Ingeniería social y educación*, I Congreso Iberoamericano de Historia de la Educación, Bogotá.

Dachary, A. & Burne (1995), S. "Educación en el Caribe" en Puiggrós, Adriana & Lozano, Claudio(Comps.), Historia de la Educación en Iberoamérica, Miño y Dávila, Buenos Aires.

Díaz, Angel (1989) "Debate en relación a la investigación curricular en

México," Furlán, Alfredo & Pasillas, Migel (comps.) *Desarrollo de la investigación en el campo del curriculum*, UNAM, México.

Díaz Villa, Mario (1993) *El campo intelectual de la educación en Colombia*, Universidad del Valle, Cali.

Ducoing, Patricia (1990) *La pedagogía en la Universidad de México 1881-1954*, CESU-UNAM.

García Guadilla, Carmen (1998) La investigación sobre la Universidad latinoamericana desde la segunda mitad del Siglo XX en Téllez, Magaldy, Educación, cultura y política. *Ensayos para la comprensión de la historia de la educación en América Latina*, Universidad Central de Venbezuela, Caracas.

Gonçalves Vidal (1996), "Arte pratica ou ciencia aplicada: o dsicurso pedagogico e a formaçao docente" en Gvirtz, Silvina (comp.) *Escuela Nueva en Argentina y Brasil. Visiones comparadas*, Miño y Dávila, Buenos Aires.

Labarca, Gullermo y otros (1977) *La educación burguesa*, Nueva Visión, México.

Levenberg, Rubén (1995) "Educación y política en Centroamérica" en Puiggrós, Adriana & Lozano, Claudio(Comps.), *Historia de la Educación en Iberoamérica*, Miño y Dávila, Buenos Aires.

Martínez Boom, Alberto & Narodowski, Mariano (1996) *Escuela, historia y poder. Miradas desde América Latina*, Ediciones Novedades Educativas, Buenos Aires.

Narodowski (1997) "Del pedagogo de Estado al pedagogo de la diversidad" *Propuesta Educativa*, Nro. 17.

Rockwell, Elsie (1991), . Ethnography and Critical Knowledge of Education in Latin America. Prospects; v21 n2.

Sánchez Gamboa, Silvio(1996) *As tendencias teorico- metodologicas nos congressos ibero-americanos de historia da educaçao*, III Congreso Iberoamericano de Historia de la Educación, Universidad Central de Venezuela, Caracas.

Tedesco, Juan Carlos (1987). Paradigms of Socioeducational Research in Latin America. *Comparative Education Review*; v31 n4.

Tellez, Magaldy (1996) "Miradas en torno a la historia de la educación latinoamericana" III Congreso Iberoamericano de Historia de la Educación, Universidad Caentral de Venezuela, Caracas.

Torres, Carlos Alberto; Puiggros, Adriana (1995). The State and Public Education in Latin America. Guest Editorial Essay.

Comparative Education Review; v39 n1.

Vasconi, Tomás(1988) *Contra la escuela. Lucha de clases y aparatos educativos en el desarrollo de América Latina,* Cuadernos de Educación, Caracas.

About the Author

Mariano Narodowski is Full Professor at the Universidad Nacional de Quilmes, Argentina. His principal areas of interest are the history of schooling processes and trends in educational policies.

Correspondence:
Universidad Nacional de Quilmes, Roque Sáenz Peña 180, (1876)
Bernal, Argentina.

E-mail: mnaro@unq.edu.ar

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 1881 times since February 8, 1999

Education Policy Analysis Archives

Volume 7 Number 3 February 8, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ
85287-0211

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Effects on Students of a 4 X 4 Junior High School Block Scheduling Program

R. Brian Cobb
Stacey Abate
Research and Development Center
for the Advancement of Student Learning
Colorado State University

Dennis Baker
Poudre (CO) School District

Abstract

The effects of a 4 X 4 block scheduling program in a middle school on a variety of student measures were investigated. These measures included standardized achievement tests in mathematics, reading, and writing, cumulative and semester grades in middle school and high school, attendance rates, and enrollment rates in advanced high school courses (in mathematics only). The block scheduling program had been in effect for four years allowing analyses of current middle and high school students who had experienced a minimum of one and one-half years of block scheduling while in middle school. The primary research design was a post-test only, matched pairs design. Students were matched on school characteristics, gender, ethnicity,

grade level, and 5th grade standardized reading scores. Results were relatively consistent with the extant literature and generally positive.

Introduction

With the advent of the public school reform movement in the 1980s, schools and school districts were barraged with criticisms and demands for educational reforms. Murphy (1990) categorized the variety of these criticisms driving the educational reform agendas into three major groups: (a) Macro-level conditions, such as the failure of the United States to maintain its competitive edge in a global marketplace, (b) school outcomes, such as declining student achievement or increasing dropout rates, and (c) school conditions, such as lack of adequate standards for students or poor quality or commitment by staff.

Responses to these various criticisms also have been clustered, for example, in federal initiatives, state mandates and policies, and local efforts as school improvements (Firestone, 1990). Often, reform initiatives are the result of interactions between two or more sources of these initiatives (c.f. Wills & Peterson, 1992; Odden & Marsh, 1990), particularly if those reforms originate at state or federal levels. The systemic reform initiatives of the 1990s, for example the school-to-work initiative (Agency for Educational Development, 1995), originated with federal legislation but impacts both state-level and ultimately local-level schooling (Fuhrman & Massell, 1992; Goertz, Floden, & O'Day, 1995). Reforms that originate at the local level, however, can be driven by levels higher up or stand- alone efforts. Block scheduling and school-based management, for example, are two such stand-alone reform initiatives whose locus has been strictly from grass roots level. While school-based management has had a relatively robust examination in the literature in recent years (Wohlstetter, Smyer, & Mohrman, 1994; Center on Educational Governance, 1995), the literature on block scheduling remains relatively scant and underpowered.

Although the variations of block scheduling are endless and idiosyncratic to the schools that implement them, all forms of block scheduling carry one common feature-- extended classroom periods of time beyond the traditional 50-minute class period. Although block scheduling has been in existence and reported in the contemporary literature since the late 1960s, it gained momentum in the late 1980s as a viable scheduling model in response to the literature on cognition supporting deeper learning by students through sustained and uninterrupted interactions with their subject matter. Recently Cawelti (1994) estimated that nearly 40% of American high schools had implemented or intended to implement some form of block scheduling, attesting to its popularity as a flexible scheduling option.

The purpose of this research was to add to the literature base on block scheduling by combining several advantageous features of research on educational innovations in general which are not typical of the block scheduling empirical literature base. These were the use of multiple measures of student effects, the use of a high-quality matched control group sampling design, and the use of a school in which block scheduling had been in place for several years.

Models of Block Scheduling

Although idiosyncratic modifications to any block scheduling model are typically implemented at any school using block scheduling, there are five general models of block scheduling that appear in the literature. One of these -- parallel block scheduling -- is used exclusively at the elementary school level and thus will not be further mentioned in this article. The other four models are used exclusively at the middle school, high school, and postsecondary levels. These are the 4 X 4 Semester Plan, the Alternative Day Plan, the Trimester Plan, and the Extended-Time Plan. Table 1 gives a brief description of these models and includes any alternative names for these models that have appeared in the recent literature.

Effects of Block Scheduling

Interest in extended periods of classroom time beyond the traditional 50-minute period first appeared in the contemporary literature under the concept of "modular scheduling", "flexible scheduling", or "modular flexible scheduling" (Polos, 1969; Stewart & Shank, 1971; Thomson, 1971; Wood, 1970). These descriptive articles arguably derived from what is considered the progenitor of block scheduling -- Carroll's (1963) seminal treatise on the theoretical advantages of extended time in school classroom periods based upon early learning theory. Presently, some 35 years later, the descriptive literature still abounds both supporting and decrying the merits of block scheduling. Fortunately the empirical literature has gradually evolved as well, and it is this literature that will be summarized briefly below.

Within this empirical literature base it is most common to encounter research on student effects and opinions of block scheduling, teacher effects and perceptions of effects of block scheduling, and parent opinions of block scheduling. Since this study focused exclusively on student effects and opinions of block scheduling, only that literature will be reported below.

Table 1

Four General Models of Block Scheduling

Model	Alternative Names	Features	Unique Advantages and Disadvantages
4 X 4 Semester Plan	Accelerated Schedule Copernican	Students enroll in four 90-minute courses that meet every day of the week for a semester, allowing completion of four year-long co	Advantages are that teachers work with fewer students, have fewer preps, and a fresh start with new students in the middle of the year. Students have only four courses to concentrate on at any one time;

	Plan	uivalent courses in a semester. Teachers typically teach three courses each semester.	they have greater opportunities for acceleration. Disadvantages are less opportunity to give homework. There seems to be less time to complete the curriculum coverage. Courses taken in the Fall semester may not be followed by a course in the same discipline for 9 months. Year-long programs/courses such as band, orchestra, and choir can be cut short.
Alternative Day Plan	A/B Odd/Even Day1/Day2	Students and teachers meet in three-four 90-120 minute classes on alternating days for the entire year.	Advantages are that teachers have the entire year for each course, with a class intensity of 90-120 minutes per course; greater opportunity to give homework due to alternating schedule; no extended latency period between courses. Disadvantages are the unevenness of scheduling with classes alternating each week as to which are on Mondays and which are on Tuesdays can be confusing.
Trimester Plan	None	Students take two-three 120 minute classes for 60 days, along with two-three traditional-length classes for the entire year.	Advantages are that it accommodates those programs/courses such as band, choir, and orchestra that need year-long contact with students, while maintaining weekly intensity of 4 X 4 Semester Plan Disadvantages are similar to 4 X 4 Semester Plan
Extended-Time Plan	Reconfiguring the Year	Schools usually partition their school year into three segments, generally including two 75-day blocks, and one 30-day block (sometimes between the 75-day blocks, sometimes at the end of the year). Then during the 75-day blocks, students enroll in three-four 90-120 minute courses daily. During the 30-day segment, students work in concentrated remediation or enrichment activities.	Advantages are that there is more flexibility inherent in the model. Disadvantages are similar to the 4 X 4 Semester Plan

The earliest empirical studies published under any of the model names described in Table 1 appeared in the late 1960s and early 1970s. The first (Steagall, 1968) compared outcomes in high school business education programs across the state of Ohio for students enrolled in 18 "block-in- time" schools with students enrolled in 18 "conventional" demographically matched schools. After adjusting for I. Q. differences in the students from both groups, results showed no significant main-effect differences in both knowledge and performance test scores as measured by the National Business Entrance Tests. When adding a three-level ability factor and a three-level urbanicity factor, only two significant results emerged -- one favoring block scheduling for urban students and one favoring conventional scheduling for suburban students.

Slightly more negative effects of block scheduling were found by Van Mondfrans, Schott, & French (1972) in an experimental study of block scheduling on performance on English tests and student attitudes toward school. Overall, conventional format students performed significantly better on the English tests and no significant differences were found on student attitudes. One consistent interaction effect was found favoring block scheduling with senior-level students over freshman, sophomores, and juniors. The design of this study seems somewhat flawed, however, since all 12 teachers participating in this study taught both traditional format and block scheduling classes on an alternating basis each day.

A decade later Sigurdson (1981; 1982) conducted two studies of the same junior high school in Canada. While his 1981 study showed little differences between block student's achievement compared to traditionally-schooled students, his second year study (Sigurdson, 1982) showed dramatically more positive results in favor of block scheduling. One conclusion that has been echoed repeatedly in subsequent studies pointed to the necessity of waiting at least a year or more. For example, Schroth and Dixon (1995) and Meadows (1995) both advocated strongly that at least three to five years of experience with block scheduling should occur in schools before valid and justifiable judgments should be drawn about effects on students.

It was not until the mid-1990s that a resurgence of empirical studies were published on block scheduling -- probably due at least in part to the exponential increase in interest -- what Shortt and Thayer (1997) call a rediscovery and redefinition. Again, many of these studies focused on teacher, administrator, and/or parent effects or opinions (i.e. Davis-Wiley, George, & Cozart; Hurley, 1997). Our interest in this review is to focus on only student effects, however, so those studies will be omitted from this review.

A number of studies have been published recently which focused on both generalized student effects (Cox, 1994; Guskey, & Kifer, 1995; Hurley, 1997; Mistretta & Polansky, 1997; Queen, Algozzine, & Eaddy, 1997; Meadows, 1995) and discipline-specific effects (Queen, Algozzine, & Eaddy, 1996; Reid, 1995; Schroth & Dixon, 1995; Wronkovich, Hess, & Robinson, 1997) of block scheduling. The results, on balance, were generally positive, but the negative findings which were reported cannot be overlooked. On the positive side, the most consistent findings that were reported were students' favorable opinions of block scheduling, particularly with teachers who found it easy to mix lecture and group-work instruction. Students also liked the fact that block scheduling seemed to reduce homework loads, although this finding would be construed as negative from other perspectives.

Beyond these kinds of qualitative student opinion effects, the findings on student achievement, attendance, and behavior/disruptions/suspensions were more equivocal. Reductions in behavior problems appeared to be relatively consistent, as were increases in attendance rates, yet if a student missed a sequence of classes for any reason, it appeared more difficult to catch up with the content and make-up assignments. At-risk students in block scheduling appeared to benefit the most consistently across the curriculum, but standardized scores on mathematics examinations were consistently lower with block scheduling.

This study adds to the literature base in several important ways. First, the great majority of empirical studies focused on the high school level, while this study reports effects on junior high school students. Second, most studies reported effects from one to two years of operating block scheduling, while the present study was implemented after four years of operation. Finally, many studies used non-probability sampling designs -- typically convenience or cluster sampling processes, while this study used a matched sampling design. None of the currently reported studies at the junior high or middle school levels used both a high quality sampling design and analysis after multiple years of operation.

Method

To implement this research, a between-groups (matched control group) design was used, with some variations depending on the particular hypothesis to be addressed. Below, each of the research hypotheses is elaborated followed by a specification of the dependent variable(s) and the specific sampling procedures for the hypothesis.

Hypotheses

This evaluation looked at effects of block scheduling on two groups of students -- junior high school students and beginning senior high school students -- along four major dependent variables -- grade point average, standardized achievement scores, attendance rates, and preparation for advanced coursework. The major research question guiding the study was: "What are the effects of block scheduling on a variety of student outcome variables?" The six hypotheses associated with this research question were:

1. Hypothesis 1 (H1): Junior high school students who experience block scheduling will evidence significantly higher grade point averages than their matched control group counterparts.
2. Hypothesis 2 (H2): Senior high school students who experienced block scheduling in junior high school will evidence significantly higher grade point averages than their matched control group

counterparts.

3. Hypothesis 3 (H3): Junior high school students who experience block scheduling will evidence significantly higher standardized test scores than their matched control group counterparts.
4. Hypothesis 4 (H4): Senior high school students who experienced block scheduling in junior high school will evidence significantly higher standardized test scores than their matched control group counterparts.
5. Hypothesis 5 (H5): Attendance rates at Block Junior High School during the years in which they experienced block scheduling will not differ significantly from their attendance rates prior to block scheduling and will not differ significantly from same-year attendance rates at matched control junior high schools.
6. Hypothesis 6 (H6): Senior high school students who experienced block scheduling in junior high school will enroll in advanced coursework at significantly higher rates than their matched control group counterparts.

Table 2 presents a matrix delineating these dependent variables along with the hypotheses that were examined for each of the two groups of students.

Table 2

Dependent Variables and Associated Research Hypotheses for Each Grade Level

Dependent Variable	Junior High	Senior High
Grade Point Average	H_1	H_2
Standardized Achievement Test Scores	H_3	H_4
Attendance Rates	H_5	*
Preparation for Advanced Coursework	*	H_6

Dependent Variables and Sampling Procedures

Hypotheses 1 and 3. The grade point average dependent variable for the Block Junior High School students and their matched controls was their Fall 1996 cumulative grade point average as maintained in the school district student data system. The standardized achievement test dependent variable was represented by three scores from the Iowa Test of Basic Skills (TAP version). These scores were the reading, mathematics, and written expression raw scores.

These two research hypotheses involved the use of a matched control group sampling design. Block Junior High School students who were in the 8th and 9th grades in the Spring of 1997, and who had experienced a minimum of three semesters of block scheduling by January 1997, were designated as the experimental group. Each of these students was then matched with students from two other junior high schools in the same geographic quadrant of the city whose school size, ethnic, and socio-economic make-up were comparable. The matching process was conducted in two stages. First, each experimental group student was matched on three demographic attributes, each with two levels: grade level (8th or 9th), gender (female or male), and ethnicity (white [non-Hispanic] or other). Once a pool of potential matched control students were identified based on these criteria, a second level of matching occurred using 5th grade Iowa Test of Basic Skills (Reading) scores. Matches were thus made with the control student whose ITBS Reading score was the closest to the experimental group student.

Hypotheses 2 and 4. For Senior High School students and their matched controls, the grade point average dependent variable was their first (Fall) term grade point average in high school. Thus, for 10th graders, it was their Fall 1996 grade point average; for 11th graders, it was their Fall 1995 grade point average. The standardized achievement test dependent variable was represented by three scores from the Iowa Test of Basic Skills (TAP version). These scores are the reading, mathematics, and written expression raw scores.

The sampling design (matched control group) and matching criteria that was used for the second and fourth hypotheses was essentially the same as was used for the first and third hypotheses above, with one key difference -- experimental group participants were 10th and 11th grade Senior High School students who had earlier experienced at least three semesters of block scheduling while attending Block Junior High School. Their respective matched controls could have attended junior high school at any of the districts' six junior high schools which were not under block scheduling.

Hypothesis 5. The dependent variable of attendance rates was the annual school-reported average daily attendance rates as reported to the Colorado Department of Education. Block Junior High School began its block scheduling in the Fall of 1993. Annual school-reported rates for attendance were obtained from the previous four years (1989-90 to 1992-93) without block scheduling and the first three active years of block scheduling (1993-94 to 1995-96). Similar rates for attendance of three comparable junior high schools were also sampled for comparison purposes.

Hypothesis 6. The dependent variable for this hypothesis was a ranking of the relative difficulty level of the mathematics course for which entering 10th graders had registered upon entry into high school. Registration rate for mathematics coursework was selected to represent the construct of "registration for advanced coursework" for two reasons. First, mathematics is a curricular area where almost every student registers during their first semester in senior high school. Second, while other areas such as science, English, social studies, and foreign languages are also high frequency registration areas, it is far more difficult to judge consistently what is advanced and what is not in these curricular areas. In mathematics, it is relatively easy to rank order the difficulty of classes in order to determine consistently advanced registration rates. Table 3 delineates the names of the various mathematics course options for which entering 10th grade students could enroll and the rankings for those course options as verified by mathematics teachers and counselors in the high school. The sampling design for this hypothesis was the same sampling design as that for H2 and H4 with one additional step. From the experimental group roster of students, those participants who did not sign up for a mathematics course (and their matched control counterparts) were eliminated from the sample.

Table 3

Name and Ranking of Mathematics Courses

Ranking	Math Course(s) Title(s)
1	Math Concepts
2	PTC Math - (remedial math for alternative placement students)
3	Algebraic Concepts
4	Applied Math 1
5	Integrated Algebra IA
6	PTC Algebra
7	Accounting I
8	Algebra I Saxon Algebra I - (taught in a different manner than traditionally) Integrated Algebra I - (taught in a different manner than traditionally)
9	Geometry Saxon Algebra II Integrated Geometry - (taught in a different manner than traditionally)
10	Algebra II Pre-International Baccalaureate Math
11	Pre-Calculus International Baccalaureate Math

Results

The results of this program evaluation are given in two major sections. First, a demographic profile for each of the samples is provided, to the extent that demographic data were collected on them. Second, the six research hypotheses are examined with statistical tables and charts provided as appropriate.

Demographic Profiles

Table 4 presents demographic and sampling data for those students who were included in the analyses associated with grade point averages and standardized achievement scores.

Table 4

Demographic Information on student samples for Grade Point Averages and Standardized Achievement Scores

				Gender		Ethnicity				
Grade Level	n	Mean 5th Gr Read.-Block	Mean 5th Gr Read.-Trad	# Girls	# Boys	#[a] As-Am	#[b] Lat.	#[c] Af-Am	#[d] Na-Am	#[e] Cauc
Eighth	96	59.34	59.38	38	58	1	5	1	-	89
Ninth	109	61.03	61.11	56	53	2	6	-	2	99
Tenth	88	59.73	59.82	50	38	2	5	-	-	81
Eleventh	62	62.11	62.19	30	32	1	2	-	-	59
TOTALS	355	60.55	60.63	174	181	6	18	1	2	328

Note.
a - Asian-American or Pacific Islander
b - Latino/Hispanic/Mexican American
c - African-American
d - Native American or Alaskan
e - Caucasian

For each of the 355 block scheduling students, a perfect match was located relative to grade level, gender, and ethnicity. For the final matching criterion, 5th grade standardized test score in reading, it was not possible in all cases to locate a matched control with exactly the same score. However, no experimental/control matched pair differed by more than 0.09 points. In addition, boys and girls were relatively equally distributed across the grades. The samples were overwhelmingly made up of Caucasian students, which is reflective both of the community and the neighborhoods in which the schools were located.

Research Hypotheses

The analyses for Hypotheses 1-4, which examined the effects of block scheduling on student grade point average (both semester and cumulative) and standardized test scores (mathematics, reading, and writing), were completed using five ANOVA's with repeated measures on the matching variable. Student gender (two levels) and student grade level (four levels) represented the between groups variables, and the block schedule students scores compared with the matched control group students scores represented the third within subjects variable. Table 5 presents the results of these five ANOVA's in abbreviated form.

As can be seen in Table 5, significant main effects for the matching variable (at $p. < .10$) existed for both Experimental/Control contrasts and for the mathematics achievement test scores. Significant first order interactions (at $p < .05$) existed for the semester GPA contrast. All other effects were not significant.

Table 5

ANOVA Results for GPA and Achievement Test Contrasts

Criterion Variablep. value			
Semester GPA	Exp/Cont*	2.75	.098
	Exp/Cont x Gender ***	7.51	.006
	Exp/Cont x Grade Level ***	4.26	.006
	Exp/Cont x Gender x Grade Level	0.67	.568
Cumulative GPA	Exp/Cont **	5.36	.022
	Exp/Cont x Gender	0.59	.441
	Exp/Cont x Grade Level	0.06	.804
	Exp/Cont x Gender x Grade Level	1.28	.259
Standardized Math	Test *	3.47	.064
	Test x Gender	0.27	.761
	Test x Grade Level	0.03	.873
	Test x Gender x Grade Level	1.70	.184
Standardized Reading	Test	0.01	.915
	Test x Gender	1.36	.246
	Test x Grade Level	0.37	.545
	Test x Gender x Grade Level	0.15	.697
Standardized Writing	Test	0.49	.483
	Test x Gender	1.68	.197
	Test x Grade Level	0.65	.422
	Test x Gender x Grade Level	0.03	.863

$p < .10$

**$p < .05$

***$p < .01$

Table 6 presents the sample sizes, means and standard deviations (for the significant within or between group contrasts only), and corresponding effect sizes for those contrasts in Table 5 that proved statistically significant at the $p < .10$ level. As can be seen from Table 6, the four significant Experimental/Control group contrasts all favored block scheduling, whereas the single significant standardized test contrast (in math) favored traditional scheduling. It is also evident that the magnitude of the significant effects, particularly those with $p < .05$, ranges relatively consistently between one quarter to one third of a standard deviation.

Table 6

Descriptive Results for GPA and Achievement Test Effects

Criterion Variable	Effects	n pairs	Mean (s.d.) Block Schedule	Mean (s.d.) Trad. Schedule	ES
Semester GPA	Exp/Cont*	346	2.91(.87)	2.82(1.00)	0.09
	Exp/Cont x Gender***	177	2.80(.92)	2.56(1.04)	0.22
	(males)				
		87	2.82(.95)	2.50(1.06)	0.30
	Exp/Cont x Grade Level***	61	2.82(.82)	2.57(0.96)	0.25
	(10th graders)				
	(11th graders)				
Cumulative GPA	Exp/Cont**	150	2.98(.76)	2.80(0.74)	0.24
Achievement Test	Math*	236	61.63(26.18)	65.36(25.24)	-0.15

Finally, Table 7 gives all of the directions of the mean differences for each of the five sets of contrasts regardless of their statistical significance.

As can be seen in Table 7, the totals are exactly equal in terms of the number of instances where the block scheduling means were greater than and less than the traditional scheduling means. It should be remembered, however, that only three of the contrasts were statistically different from each other at the $p < .05$ level, and all three favored the block scheduling students.

Table 7

Numbers of Contrasts in which the Mean of the Block Scheduling Students is Greater than and Less than the Mean of the Traditional Scheduling Students

Criterion Variable	Effects (# of possible contrasts)	# contrasts when mean block > trad.	# contrasts when mean trad. > block
Semester GPA	Exp/Cont (1)	1	0
	Exp/Cont x Gender (2)	1	1
	Exp/Cont x Grade Level (4)	3	1
Cumulative GPA	Exp/Cont (1)	1	0
	Exp/Cont x Gender (2)	2	0
	Exp/Cont x Grade Level (2)	2	0
Standardized Math	Test (1)	0	1
	Test x Gender (2)	0	2
	Test x Grade Level (3)	0	3
Standardized Reading	Test (1)	1	0
	Test x Gender (2)	0	2
	Test x Grade Level (2)	1	1
Standardized Writing	Test (1)	0	1
	Test x Gender (2)	1	1
	Test x Grade Level (2)	1	1
TOTALS	28 possible contrasts	14	14

Hypothesis 5 focused on an examination of comparative attendance rates at Block Junior High School before and after initiating block scheduling, and with three comparable schools. Figure 1 displays the average daily attendance rates of all four schools for the four years prior and three years after initiating block scheduling. As can be seen in Figure 1, Block Junior High School's ADA rate had been declining slightly each year (from a 94.5 attendance rate in 1989-90 to a 93.7 rate in 1992-93) prior to implementing block scheduling. Attendance rates reversed this trend beginning in the Fall of 1993 and climbed consistently about 0.1 percentage point each of the three years of block scheduling. Figure 1 does not show any clear patterns of attendance rates for the other three junior high schools during the four pre-block scheduling and three post-block scheduling years.

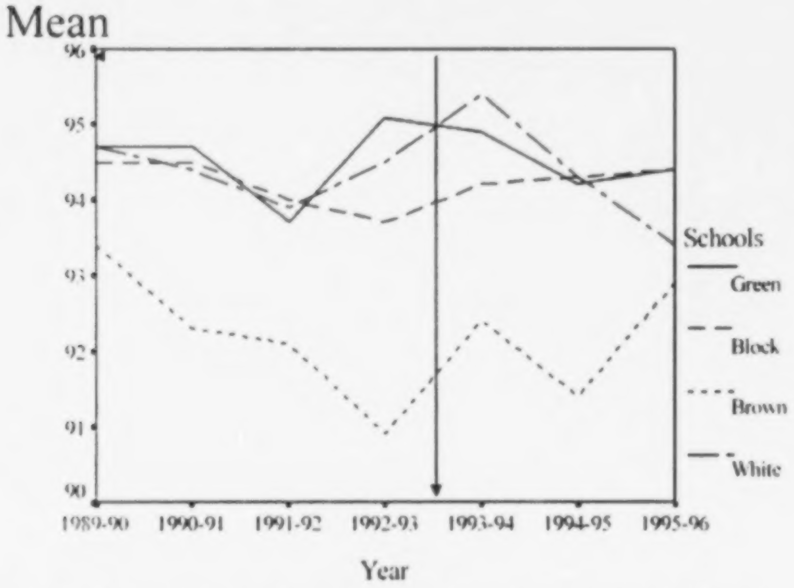


Figure 1. Comparison of average daily attendance rates of Block Junior High School with rates at three comparable junior high schools with traditional scheduling. (The vertical line in the center of the figure signifies the point at which block scheduling was implemented at Block Junior High School.)

Hypothesis 6 was concerned with comparing advanced coursework registration rates in 10th grade mathematics by students who had experienced block scheduling versus those who had not. Referring back to Table 3, it can be seen that there were eleven possible mathematics courses into which students could register at the high school to which they matriculated upon completion of their junior high school experience. Table 8 presents the results of a Wilcoxin signed ranks matched pairs test. As can be seen, the mean ranked mathematics course registration was virtually identical for both block scheduled and traditionally scheduled students.

Table 8

t-test of Advanced Mathematics Course Registration

	Descriptive Statistics			Inferential Statistics	
Comparison Group	Mean	N	Std. Dev.	T-Value	P-Value
Junior High School Students	8.57	238	2.12	0.189	.851
Comparison School Students	8.54	238	1.83		

Discussion

In prior literature, there were only three empirical studies examining student effects of block scheduling at the middle/junior high school level. Two of them were sequential examinations of a variety of student effects at the same junior high school (Sigurdson, 1981; Sigurdson, 1982) in Alberta, Canada; Schroth and Dixon (1995) looked strictly at mathematics achievement effects of block scheduling on students at two middle schools in Texas. In both locations, a 4 X 4 schedule was used; none of the three looked at effects past two years of implementation, and both sets of authors indicated that more time was necessary in order to ascertain true effects of this scheduling intervention.

Schroth and Dixon (1995) did find, however, that mathematics achievement did not differ significantly between those students prepared using block scheduling compared with those prepared under traditional scheduling. Sigurdson (1982) also found no differences between block-scheduled and traditionally-scheduled students in their attitudes towards mathematics nor in their mathematics achievement scores. Similar "no differences" findings relative to mathematics effects were found by Lockwood (1995) and the North Carolina Department of Public Instruction (1995), although these studies were both conducted at the high school level.

The findings of this study confirm the "no differences" conclusion in registration for advanced mathematics courses, but conflict with the "no differences" finding on mathematics achievement. The block-scheduled students in this study performed significantly less well on standardized mathematics tests compared with their traditionally- scheduled peers. The "no differences" findings in this study in standardized reading and writing test scores also is consistent with the findings of Holmberg (1996). Again, Holmberg was studying standardized test scores of high school students rather than junior high school students.

On the positive side, this study found consistently higher grade point averages, both semester and cumulative, in favor of block scheduled students. These kinds of non- standardized achievement effects have been reported in the high school literature on block scheduling (Buckman, King, & Ryan, 1995; Reid, Hierck, & Veregin, 1994; Payne & Jordan, 1996), but have been contradicted by Parkinson and Parkinson (1995) at the postsecondary level.

At a more complex level this study found statistically significant interactions suggesting block scheduling has a more positive semester GPA effect on male students compared with female students and for 10th and 11th graders compared with 8th and 9th graders. These interactions did not hold for cumulative GPA. No other studies specifically examined differential effects by grade level, and Lockwood's (1995) study was the only study reporting out gender effects, which were not found to be present.

Finally, this study examined attendance data. This variable was a popular one in the literature with positive findings at the high school level reported by Buckman et. al., (1995) and Reid et. al., (1994), and no differences reported by Guskey and Kifer (1995). This study did not test for statistically significant differences, but did provide a visual examination of attendance rates over time and across several schools. The findings here are consistent with those cited above.

Very little that is definitive can be inferred from this study. As mentioned earlier, its most positive contributions would be that it begins to fill a significant void in the middle/junior high school literature on effects of block scheduling. This contribution is made all the more important given the relatively high-quality matched group design, and given that block scheduling had been in effect for four years when the data for this study were collected.

As is usually the case, this study points to as many questions as answers, however. In addition to the necessity for more high-quality studies at the middle/junior high school level, two very fertile areas of follow-up research were suggested from this study. First, variables of gender and ethnicity need to be embedded in every design in future studies to clarify as much as possible differential effects. If the gender or effects found in this study, or other kinds of attributional effects hold up over time, these findings may present the most damaging or even fatal flaws in block scheduling reported to date.

Just as important is the need for researchers to ratchet up the "block versus traditional schedule" contrast variable to account for multiple ways of implementing block scheduling. There are sufficient schools, for example, which are implementing alternate day scheduling now to allow for multiple block scheduling contrasts to be made instead of the typical single contrast between block scheduling and traditional scheduling. Similarly, there are now enough modifications that have been reported to 4 X 4 semester block scheduling to allow for multiple contrasts even within this single model of block scheduling.

This research venue may ultimately provide the most illuminating information about the potential for block scheduling in the schools. Many 4 X 4 semester block scheduled schools are currently leaving room in the schedule for year-long, 50-minute classes which can accommodate the needs for year-long attention to mathematics, band, chorus, and advanced placement classes. These kinds of modifications probably hold the key in the long run to establishing the flexibility in scheduling to make the best use of the characteristics educators like of both traditional and block scheduling.

References

Agency for Educational Development, (1995). *Study of school-to- work reform initiatives. Volume I: Findings and Conclusions*. Washingt on, DC: Author.

Buckman, D., King, B., & Ryan, S. (1995). Block scheduling: A means to improve school climate. *NASSP Bulletin, 79*(571), 9-18.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*, 723-733.

Cawelti, G. (1994). *High school restructuring: A national survey.* Arlington, VA: Educational Research Service.

Center on Educational Governance, (1995). The school-based management study. Volume I: Findings and conclusions. Los Angeles, CA: University of Southern California, School of Education.

Cox, L. (1994). A study of the effects of a block scheduling program with high school students who are "at-risk". Unpublished doctoral dissertation, University of Houston.

Davis-Wiley, P., George, M., & Cozart, A. (1995, November). Block scheduling in the secondary arena: Perceptions from the inside. Paper presented at the annual conference of the Mid-South Educational Research Ass ociation, Biloxi, MS.

Firestone, W. A. (1990). Continuity and incrementalism after all: State responses to the excellence movement. In J. Murphy (Ed.). *The educat ional reform movement of the 1980s: Perspectives and cases.* Berkeley, CA: McCutchan.

Fuhrman, S. H., & Massell, D. (1992). Issues and strategies in systemic reform. New Brunswick, NJ: Rutgers University, Consortium for Policy Research in Education.

Goertz, M. E., Floden, R. E., & O"Day, J. (1995). Studies of education reform: Systemic reform. Volume I: Findings and conclusions. New Brunswick, NJ: Rutgers University, Consortium for Policy Research in Education.

Guskey, T. R., & Kifer, E. (1995, April). Evaluation of a high school block schedule restructuring program. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Holmberg, T. (1996). Block scheduling versus traditional education: A comparison of grade-point averages and ACT scores. Unpublished doctoral dissertation, University of Wisconsin, Eau Claire.

Hurley, J. C. (1997b). The 4 x 4 block scheduling model: What do students have to say about it? *NASSP Bulletin, 81*(593), 64-72.

Lockwood, S. (1995). Semesterizing the high school schedule: The impact on student achievement in algebra and geometry in Dothan City Schools, Dothan, Alabama. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa.

Meadows, M. E. (1995). A preliminary program review of the four-period day as implemented in four high schools. Unpublished doctoral dissertation, University of Maryland, College Park.

Mistretta, G. M., & Polansky, H. B. (1997). Prisoners of time: Implementing a block schedule in a high school. *NASSP Bulletin, 81*, (593), 23-31.

Murphy, J. (1990). The educational reform movement of the 1980s: A comprehensive analysis. In J. Murphy (Ed.). *The educational reform movement of the 1980s: Perspectives and cases.* Berkeley, CA: McCutchan.

North Carolina Department of Public Instruction. (199). Block scheduled high school achievement: Comparison of 1995 end-of-course test scores for blocked and non-Blocked high schools [On-line]. Available: http://www.dpi.state.nc.us/block_ block_scheduling_report/#4

Parkinson, S., & Parkinson, C. (1995). Examination scores: Block teaching versus traditional teaching. *Nurse Educator, 20*(4), 14.

Payne, D., & Jordan, M. (1996). The evaluation of a high school block schedule: Convergence of teacher and student data. *American Secondary Education, 25*(2), 16-19.

Queen, J. A., Algozzine, B., & Eaddy, M. (1996). The success of 4 x 4 block scheduling in the social studies. *The Social Studies, 87* (6), 249-253.

Queen, J. A., Algozzine, R. F., & Eaddy, M. A., (1997). The road we traveled: Scheduling in the 4 x 4 block. *NASSP Bulletin, 81* (588), 88-99.

Odden, A., & Marsh, D. (1990). Local response to the 1980s state education reforms: New patterns of local and state interactions. In J. Murphy (Ed.). *The educational reform movement of the 1980s: Perspectives and cases.* Berkeley, CA: McCutchan.

Polos, N. C. (1969). Flexible scheduling: Advantages and disadvantages. *Education, 89,* 315-319.

Reid, L. (1995). Perceived effects of block scheduling on the teaching of

English. (ERIC Document Reproduction Service No. ED 382 950).

Reid, W., Hierck, T., & Veregin, L. (1994). Measurable gains of block scheduling. *School Administrator, 51*, 32-33.

Shortt, T. L., & Thayer, Y. V. (1997). A vision for block scheduling: Where are we now? Where are we going? *NASSP Bulletin, 81*(593), 1-15.

Schroth, G., & Dixon, J. (1995). The effects of block scheduling on seventh grade math students. East Texas State University. (ERIC Document Reproduction Service No. ED 387 887)

Sigurdson, S. (1981). The block plan. An alternative approach to the needs of junior high school students._Final report, 1981. Alberta, Canada. (ERIC Document Reproduction Service No. ED 221 623.

Sigurdson, S. (1982). Two years on the block plan. Meeting the needs of junior high school students. Final report, 1982. Alberta, Canada. (ERIC Document Reproduction Service No. ED 225 946.

Steagall, P. H. (1968). A study of the bolck-of-time scheduling in the secondary business and office education program in Ohio. Unpublished doctoral dissertation, The Ohio State University, Columbus.

Stewart, J. W., & Shank, J. (1971). Daily demand of modular flexible scheduling for small schools. *Educational Leadership, 29*, 537-544.

Van Mondfrans, A. P., Schott, J. L., & French, D. G. (1972). Comparing block scheduling and traditional scheduling on student achievement and attitudes. Paper presented at the annual convention of the American Educational Research Association, Chicago, IL.

Wills, F. G., & Peterson, K. D. (1992). External pressures for reform and strategy formation at the district level: Superintendent's interpretations of state demands. *Educational Evaluation and Policy Analysis, 14*(3), 241-260.

Wohlstetter, P., Smyer, R., & Mohrman, S. A. (1994). New boundaries for school-based management: The high involvement model. *Educational Evaluation and Policy Analysis, 16*(3),

Wood, C. L. (1970). Modular scheduling? Yes but ---. *Journal of Secondary Education, 45*, 40-42.

Wronkovich, M., Hess, C. A., & Robinson, J. E. (1997). An objective look a math outcomes based on new research into block scheduling. *NASSP Bulletin, 81*(593), 33-41.

About the Authors

Brian Cobb

School of Education
Colorado State University

Email: Cobb@CAHS.Colostate.edu

Brian Cobb is a Professor in the School of Education at Colorado State University and Co-Director of the Research and Development Center for the Advancement of Student Learning, a community research collaborative in Ft. Collins, Colorado. His research interests presently focus on a variety of educational reform topics including charter schools, high stakes testing, and block scheduling.

Dennis Baker
Poudre (CO) School District
Ft. Collins, Colorado

Dennis Baker has been involved with education for many years as a teacher, counselor, and administrator. Since 1995, he has been a principal with the Poudre School District in Ft. Collins, CO. In the fall of 1998, he assumed his current position as principal at Fort Collins High School. Presently, he is completing his doctoral program at the University of Northern Colorado in Educational Leadership.

Stacey Abate
Poudre (CO) School District
Ft. Collins, Colorado

Stacey Abate, M.Ed., has been the Grants Coordinator for the Poudre School District for the past 2 years. She also serves as a research assistant at the Research and Development Center for the Advancement of Student Learning. Her prior educational experience includes appointments as an interpreter and teacher of the Deaf in New Mexico and Australia.

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

~ EPAA Vol. 7 No. 4: Haney, Fowler, Wheelock, Bebell & Malec "Massachusetts Teacher Test" ~

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

Interviews with MTT-Takers: Vignettes and Summary

Several individuals who sent copies of their MTT score reports to the Ad Hoc Committee spontaneously offered us comments on the new MTT. For example, one woman wrote:

> After graduation from xx College--one of the best schools in the area, ALL of my daughter's friends failed at least one section of the MA test. Is something wrong with this picture? After a phone chain among many parents, we all agree there is a problem with the MA test, as these students did all pass required testing for other states. [Underline in original; name of College deleted to protect confidentiality.]

Another correspondent wrote:

> My scores on the Praxis series earned me a license to teach Language Arts and Social studies to grades six through nine in North Carolina. Unfortunately, this was not enough to earn a reprieve from the Massachusetts test. This is just one of the aspects of the test with which I take issue.
>
> One major problem with the Massachusetts Teacher Tests is that there is no preparation offered. When I called to request information on the test, I received a packet of test objectives for each of the tests I was taking. This information was practically useless, as I still had no clue as to the format of the test. [Letter to the Ad Hoc Committee, dated August 20, 1998.]

In light of such comments, we contacted the first 15 test-takers who had sent us copies of their score reports, to ask whether they were be willing to be interviewed, on condition that we keep their identities confidential. All 15 agreed. The interviews focused on the current professional status of the teacher candidates, sought their views on the administration and content of the MTT tests and asked about their attitudes toward testing and teaching. We gathered this information in telephone interviews lasting between one half to one hour during November 1998. We took notes during phone conversations and elaborated them after the end of conversations. Typewritten accounts of each interview were then prepared, and results across interviewees were analyzed by looking for common themes and comments.

Interview sample

Although this was a small, self-selected sample, those who agreed to an interview represented a wide range of experiences. Of the fifteen, seven (47%)

• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 2119 times since February 11, 1999

Education Policy Analysis Archives

Volume 7 Number 4 February 11, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ
85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Less Truth Than Error?
An independent study of the Massachusetts Teacher Tests

Walt Haney, Clarke Fowler, Anne Wheelock, Damian Bebell and Nicole Malec

Ad Hoc Committee to Test the Teacher Test

Center for the Study of Testing, Evaluation and Educational Policy
Boston College

Abstract

The Massachusetts Teacher Tests (MTT), introduced last year, have never been subject to external review as required by the measurement profession's standards and many legal precedents. Neither the Massachusetts Department of Education (DOE) nor the tests' manufacturer have made public information about the exams' reliability (consistency) or validity (meaningfulness). Using data from state and academic reports from the April and July test dates, an ad hoc committee of nationally-known researchers has now been able to make a preliminary assessment of the exams. The committee focused on the Communications and Literacy exam that was required of all prospective teachers regardless of grade level or subject area. The

purpose of the analysis was to determine the accuracy of the tests in assessing the reading and writing skills of the test-takers.

Scores on the Massachusetts Teacher Tests of reading and writing are highly unreliable. The tests' margin of error is close to double to triple the range found on well-developed tests. A person retaking the MTT several times could have huge fluctuations in their scores even if their skill level did not change significantly. In fact, the 9 to 17 point margin of error calculated for the tests represents more than 10 percent of the grading scale (assumed to be 0 to 100). The large margin of error means there is both a high false-pass rate and a high false-failure rate. For example, a person who received a score of 72 on the writing test could have scored an 89 or a 55 simply because of the unreliability of the test. Since adults' reading and writing skills do not change a great deal over several months, this range of scores on the same test should not be possible. While this test is being touted as an accurate assessment of a person's fitness to be a teacher, one would expect the scores to accurately reflect a test-taker's verbal ability level. In addition to the large margin of error, the MTT contain questionable content that make them poor tools for measuring test-takers' reading and writing skills. The content and lack of correlation between the reading and writing scores reduces the meaningfulness, or validity, of the tests. The validity is affected not just by the content, but by a host of factors, such as the conditions under which tests were administered and how they were scored. Interviews with a small sample of test-takers confirmed published reports concerning problems with the content and administration.

If the Commonwealth wants high standards for its teaching force, it should use assessments that meet high professional standards. The current MTT fail this criterion. Results from the April and July administrations of the MTT reveal that these new tests are so unreliable and of such poor validity that they are passing candidates who should fail and failing ones who should pass. Therefore, the ad hoc committee recommends:

1. The Massachusetts Board of Education should immediately suspend the administration of the Massachusetts Teacher Tests.
2. The Commonwealth should convene an independent panel to audit the tests' development, administration, and use.
3. An investigation should be launched to uncover why the state contracted with this test developer even after learning of the company's poor past performance in developing tests of this type.

- **Introduction**
- **Background**
- **Reliability and Validity of the Mass. Teacher Tests**
- **Interviews with MTT-Takers**
- **Conclusions & Recommendations**
- **References**

Appendices

1. The Massachusetts Teacher Tests: A Chronology
2. Richardson v. Lamar County Bd. of Educ. 729 F. Supp 806 (M. D. Ala. 1989)
3. Summary of Results of Interviews with Examinees

About the Authors

Walt Haney is a professor in the School of Education and Senior Research Associate in the Center for the Study of Testing, Evaluation and Educational Policy at Boston College. He is former editor of the journal *Educational Measurement: Issues and Practice*, advisor to the committee that developed the 1985 *Standards for Educational and Psychological Testing* and author of numerous articles concerning educational testing and evaluation. He has also served as an expert witness in numerous court cases concerning testing.

Clarke Fowler is a professor in the Education Department at Salem State College. He has taught both preschool and kindergarten and currently teaches courses in early childhood education.

Anne Wheelock, an independent education policy writer and researcher, works for several national foundations and is the author of *Safe To Be Smart: Building a Culture for Standards-Based Reform in the Middle Grades* (1998).

Damian J. Bebell is a doctoral student at Boston College where he is employed at the Center for the Study of Testing, Evaluation and Educational Research. His research interests include educational philosophy, alternative forms of assessment, and homeschooling.

Electronic mail addresses and phone numbers for the authors are: haney@bc.edu (617-552-4199), clarke.fowler@salem.mass.edu (617-524-4704), wheelock@shore.net (802-254-2796), and bebell@bc.edu.

Center for the Study of Testing, Evaluation and Educational Policy
Campion Hall, Boston College, Chestnut Hill, MA 02467

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North
Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez

Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)

Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

~ EPAA Vol. 7 No. 4: Haney, Fowler, Wheelock, Bebell & Malec "Massachusetts Teacher Test" ~

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

Introduction

The Ad Hoc Committee is an independent group, whose formation is described in this report. The drafting and publication of this report have been solely the responsibility of the Committee. Nonetheless, we would like to acknowledge the support and help of numerous people and institutions in preparing this report. These include the Commonwealth Education Deans' Council, Boston College, Bridgewater State College, Elms College, Framingham State College, Lesley College, Salem State College, The University of Massachusetts at Boston, and Westfield State College. In particular, we would like to thank the Center for the Study of Testing and Public Policy, which allowed us to use its address as a temporary mailing address for the Committee. Among individuals who have generously assisted us are Irwin Blumer, Mary Brabeck, Joseph Caruso, John Cawthorne, Richard Clark, Bill Dandridge, Patricia Delaney, Anne Harrison, Virginia Harvey, Catherine Horn, Bailey Jackson, Diane Joyce, Joanne McCourt, Patricia O'Brien, Joan Rasool, Maria Sachs, Bob Schaeffer, Kelly Shasby, Dennis Shirley, and Michael Thomas. Also, we thank Larry Ludlow, Ron Hambleton and Dan Koretz who provided helpful reviews of statistical analyses recounted in this report. The report itself is entirely the responsibility of the authors, and has not been sponsored, funded or endorsed by any institutions or individuals who have aided our inquiry. Reviewers of drafts of this report have been generous in offering comments and suggestions, but naturally not all have agreed with all that is written here.

The Ad Hoc Committee was formed out of concern that important decisions were being based on Massachusetts Teacher Tests (MTT) (Note 1) scores without reasonable evidence on their reliability and validity--a clear violation of professional standards concerning testing. These standards, the widely recognized 1985 *Standards for Educational and Psychological Testing* (Note 2), have been in existence for almost 50 years (in current and previous editions) and have been relied upon in numerous legal proceedings that involved testing in state and federal courts. The very first provision of these standards deals with test validity, requiring that:

> Evidence of validity should be presented for the major types of inferences for which the use of a test is recommended. A rationale should be provided to support the particular mix of evidence presented for the intended uses. (Standard 1.1 p. 13).

Other standards call on test publishers or developers to document the reliability of test scores for each total score, subscore, or combination of scores that is reported (Standard 2.1, p. 29); to clearly describe scales used for reporting scores (Standard 4.1, p. 33); and to document the reliability of classification decisions based on licensure or certification tests (Standard 11.3, p. 65). Moreover, Standard 5.1 requires that: "A technical manual should be made available to prospective test users at the time a test is published or released for operational use" (p. 35).

In contravention of these requirements, MTT results are being used to make decisions about prospective teachers in Massachusetts, and about educational policies in the Commonwealth. Our Committee therefore set out to gather evidence on the technical merits of the MTT. Before we recount what information we have gathered and what has been learned from it, we describe the background to our inquiry.

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

~ EPAA Vol. 7 No. 4: Haney, Fowler, Wheelock, Bebell & Malec "Massachusetts Teacher Test" ~

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

Background

Massachusetts Commissioner of Education Robert Antonucci reported to the Massachusetts Board of Education on December 15, 1997, that he had selected National Evaluation Systems (NES), a test development company based in Amherst, MA, to develop and administer new teacher certification tests for Massachusetts (see appendix 1 for more details on the chronology of the development of the MTT). We learned of this in January, 1998. One of us (Haney) was sufficiently concerned about this prospect that he faxed to then Associate Commissioner of Education David Driscoll a copy of a federal court decision in which NES had been found to have "violated the minimum requirements for professional test development" when it created teacher certification tests for the state of Alabama. (Note 3) Associate Commissioner Driscoll did not respond directly but we learned through an intermediary that NES had been one of only three contractors to bid on the Request for Responses (RFR) issued on February 24, 1997. In February 1998, the Commonwealth commissioned NES to develop tests of communication and literacy skills and of subject matter knowledge. (Note 4)

At that time, the plan was that there would be two trial administrations of the new tests (in April and July, 1998) and that their scores would not count toward certification of eligible teacher candidates. As described in a study guide and registration bulletin released by the Massachusetts Department of Education (DOE) in January 1998, candidates would be required to achieve qualifying scores in order to be certified only beginning with the third administration in October 1998. But then a change was made: on March 25, 1998, the DOE announced that candidates taking the April and July tests would not qualify automatically; instead, they would have to make passing scores to be provisionally certified.

The first MTT tests were given on April 4, 1998. In the ensuing weeks NES convened scoring panels to help determine where passing scores on the new tests should be set. On June 22, acting on the recommendation of NES, the Massachusetts Board of Education voted to set passing scores one standard error of measurement below those that resulted from NES's analyses of scoring panel reviews. This meant that 44% of the 1,800 prospective teachers who took the April tests would have failed.

Then in the midst of his campaign for election as Governor, Massachusetts Acting Governor Paul Cellucci had Board of Education Chairman John Silber convene a special meeting of the Board to reconsider its vote on qualifying scores. Reversing its decision of less than two weeks earlier, the Board voted on July 2 for higher passing scores; now 59% of the April test-takers would fail. (At this meeting, Interim Commissioner of Education, Frank Haydu, appointed after Antonucci had taken a job in the private sector, resigned. Haydu had recommended the lower qualifying scores because the test was untried and subject to measurement error.)

The results of the first MTT administration were mailed to test-takers in early July. They showed that 70% of examinees passed the reading test, 59% the writing test, and widely varying percentages the 32 subject matters tests. Because candidates had to pass all three tests to pass the MTT overall, the overall passing rate was only 41%.

This overall result, with a majority of candidates failing, led to immense negative publicity. Even before the passing scores were adjusted in July, in a speech to the Greater Boston Chamber of Commerce on June 25, Massachusetts Speaker of the House Tom Finneran commented, "I'll tell you who won't be a teacher. The idiots who flunked that test and flunked so miserably--and, of course, the idiots who passed them." (As reported in the Boston Herald on 6/26/98, by Darrell S. Pressley, "Dumb struck: Finneran slams 'idiots' who failed teacher tests.")

The second MTT administration took place on July 11, just days after April test-takers had received their results. In a press release dated July 23, 1998, the DOE said:

> MALDEN--State Commissioner of Education David P. Driscoll today released the results of the April 4 Massachusetts Teacher tests showing the pass and fail rates for prospective teachers by the institution they attended.
>
> The 1795 prospective teachers who took tests in April in communication and literacy and in various subjects entered on their registration sheets information concerning the 56 colleges and universities they attended. That data was then verified by the institutions; 54 checked the data and made corrections where necessary.
>
> Because in some cases there is a very small sample of students taking specific subject tests and a small sample of graduates from some of the institutions, results need to be interpreted carefully. Many schools have fewer than ten students who took parts of the test, so making a broad statistical conclusion in those cases is not sound.

Release of results by institution prompted much public hand-wringing among teacher preparation institutions. A front page story in the Boston Globe on July 24, 1998, for example, reported, "Colleges vow to retool after failures on teacher tests" (O'Brien, 1998).

All this heightened our concern that important decisions were being made on the basis of the hastily-developed MTT before their technical merits had been established. Hence in July, at the suggestion of Clarke Fowler, the three of us (Fowler, Haney, and Wheelock) decided to found the Ad Hoc Committee to Test the Teacher Test. Our initial idea was to ask people who had taken both the MTT and some post-collegiate national examination, such as the Praxis (the successor to the National Teacher Examination or NTE) or the Graduate Record Examination (GRE), to send us copies of their score reports. Comparing the two sets of scores would allow us to test the validity of the new MTT against established tests such as the Praxis or GRE. The principle was simple: if the new MTT were valid and reliable, then scores on the new Massachusetts test should be comparable with those from other tests.

Flyers inviting test-takers to send us copies of their score reports were distributed at many of the testing sites for the July 11 administration of the MTT. As of December 1998, more than 30 individuals had generously provided us with copies of their score reports on both the MTT and some other test (including the Praxis, the Millers Analogies Test, the New York State Liberal Arts and Sciences teacher certification test, and the Texas teacher certification test, the ExCET). Among them, however, there were only twelve cases in which people had taken the MTT and the same other test, leaving us with samples that were too small for reasonable statistical analysis. Nonetheless, many individuals who sent us scores spontaneously offered us comments on the new MTT. Hence we decided to interview those willing to be interviewed. These interviews are summarized in part 4 and appendix 3 of this report.

On August 17, detailed results of the July administration were released to the institutions of higher education whose students took the tests. Seeing a copy of these results for one institution, which listed students' MTT scores from both April and July, we realized that we could use these data to estimate the reliability of the MTT tests. Adults' basic skills in reading and writing would not change much in a three-month period; people could not cram for the July test since no study guide was available; and, in any case candidates did not learn they would have to retake the test until just days before the second administration when the April results were mailed to examinees. Hence, we set out to acquire data from institutions which had students take the MTT in both April and July. The results of this inquiry are presented in part 3 of this report.

However, before describing the nature and results of this inquiry, we point out that many share our misgivings about the lack of technical documentation on the new MTT. For example, testing experts such as George Madaus of Boston College and Ron Hambleton of the University of Massachusetts at Amherst have publicly commented on the lack of documentation about the new MTT (Madaus, 1998; Hambleton, 1999); and in both press releases and letters, the Commonwealth Education Deans' Council has expressed concern about the lack of evidence on the MTT's validity (Jackson, 1998).

On August 24, 1998 Board of Higher Education Chairman James Carlin called a meeting at Framingham State College for deans and presidents of Massachusetts institutions of higher education to discuss implications of the MTT results. At this meeting, DOE spokesman

Alan Safran said: "We're committed to full disclosure about this test," and stated that the Department was willing to release all "non-proprietary" data. The Ad Hoc Committee therefore wrote to Acting Commissioner of Education David Driscoll to request three sets of data from the Massachusetts Teacher Tests that would allow us to analyze their psychometric properties. Specifically we asked for:

> 1. Item level and total test scores (both raw scores and scaled scores) for all examinees who took the Massachusetts Teacher Tests in April. We seek these data in order to be able to analyze the psychometric properties of items on the April test.
> 2. Item level and total test scores (both raw scores and scaled scores) for all examinees who took the Massachusetts Teacher Tests in July. We seek these data in order to be able to analyze the psychometric properties of items on the July test.
> 3. Test scores of all examinees (both raw scores and scaled scores) for all examinees who took the Massachusetts Teacher Tests in both April and July. We seek these data in order to analyze the test-retest reliability of the Massachusetts Teacher Tests. (Ad Hoc Committee letter to Driscoll, September 7, 1998)

Acting Commissioner Driscoll did not reply.

More recently, in a November 22, 1998, commentary article in the Boston Globe, "Good teachers need a good test," Emanuel Mason, the chair of Northeastern University's Department of Counseling Psychology, Rehabilitation, and Special Education wrote:

> The primary problem with the MTT in all its versions is its validity. Validity is the degree to which a test measures what it was designed to measure. . . . If a test does not have validity, it does not measure anything. Further, validity should be demonstrated before a test is used as the basis for decisions on licensing or any other outcome. The MTT's developer has yet to provide evidence of validity, and the test already has been administered three times.

Ignoring well-established professional standards concerning testing, the DOE and NES have not made available any documentation on the reliability and validity of the MTT. We therefore set out to study the technical merits of the tests, to begin to answer the question Mason and others have raised: How good are the new MTT tests?

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

~ EPAA Vol. 7 No. 4: Haney, Fowler, Wheelock, Bebell & Malec "Massachusetts Teacher Test" ~

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

The Reliability and Validity of the Massachusetts Teacher Tests

Given the publicity that has surrounded the new tests and the questions that have been raised about their validity and reliability, it is not surprising that Massachusetts officials have sought to defend their merits. For example, in his July 7, 1998, editorial in the New York Times, John Silber wrote that the exams had been "validated by teachers and scholars who prepared it . . . [and] again by the panels of distinguished teachers, administrators and college professors who reviewed the questions for fairness and agreed on minimal passing scores." What this defense does not take into account is that a test cannot be validated simply by having people review test questions.

Test validation refers to the meaning of test *scores* and that meaning depends not just on test content, but also on a host of other factors, such as the conditions under which tests are administered and how they are scored. A simple example illustrates this point. Suppose that we have a test made of 50 three-digit addition problems such as 231 + 458 = ? On its surface, this would seem to be a test of ability to add three-digit numbers. Perhaps so, if given in a math class with 20 or 30 minutes to solve the 50 problems. But suppose the test was sprung with little warning on aspiring accountants as a condition for getting a job, and they were given only five minutes to solve the 50 problems. Under these conditions, the test would obviously measure the ability not just to add three-digit numbers, but also to work fast under pressure. Or suppose that answers above 999 were scored correct only if they included a comma between the hundreds and thousands positions (such that 1,200 would be scored correct, but 1200 would not). If examinees were not told of this scoring rule, this would undermine the validity of the test as a measure of addition skills; the scoring rule would in effect test examinees' adherence to a particular convention for writing numbers greater than 999.

This example is directly relevant to the Massachusetts Teachers Tests, for when candidates signed up to take the April exams, they had been told that these were merely practice tests and results would not count toward certification. But less than two weeks before the examination, the DOE announced that the results *would* count toward certification. Moreover, people taking the MTT in April and July had no access to sample tests or details on how questions (such as exercises in summarization and dictation) would be scored. Hence it is impossible to assess how meaningful the MTT scores are simply by reviewing questions that make up these tests.

The concepts of test validity and reliability

So how does one assess the validity and reliability of test scores?

The 1985 test *Standards* says:

> Validity is the most important consideration in test evaluation. The concept refers to the appropriateness, meaningfulness and usefulness of the specific inferences made from test scores. Test validation is the process of accumulating evidence to support such inferences. (AERA, APA & NCME, 1985, p. 9, emphasis added)

Traditionally, three types of validity evidence have been recognized: content-related validity evidence, criterion-related validity evidence, and construct validity evidence. Content-related validity refers to the "degree to which the sample of items, tasks or questions on a test are representative of some defined universe or domain of content" (AERA, APA & NCME, 1985, p. 10). As Emanuel Mason pointed out in his November 22, 1998, article in the Boston Globe, this is the only form of validity evidence to which state and NES officials have even referred, and even here they have provided no technical documentation as required by the 1985 test *Standards*.

But, since validation refers to the meaningfulness of test scores, validation must also consider evidence on criterion-related validity and construct validity. Criterion-related evidence "demonstrates that test scores are systematically related to one or more outcome criteria" (AERA, APA & NCME, 1985, p. 11). The validity of college admissions tests is often evaluated in terms of the extent to which scores predict success in college as measured by freshman-year grade-point average (a form of criterion-related validity evidence referred to as predictive validity evidence). Another form of criterion-related validity evidence is concurrent validity. This refers to how scores on one test relate to those on another test intended to measure the same trait, when both tests are taken at about the same time. This is the sort of validity evidence that the Ad Hoc Committee was seeking when we asked test-takers to send us score reports on both the MTT and other tests designed to measure communications skills and/or teaching competence. As recounted above, we have not been able to acquire enough data to allow us to undertake a concurrent validity study.

Construct validity is an over-arching concept referring to evidence that test scores represent "a measure of the psychological characteristic of interest" (AERA, APA & NCME, 1985, p. 9):

> The process of compiling construct-related evidence for test validity starts with the process of test development and continues until the pattern of empirical relationships between test scores and other variables clearly indicates the meaning of the test score. Especially when multiple measures of a construct are available -- as in many practical testing applications -- validating inferences about a construct also requires paying careful attention to aspects of measurement such as test format, administration conditions, or language

> level, that may affect test meaning and interpretation materially.
>
> Evidence for construct interpretation of a test may be obtained from a variety of sources. Intercorrelations among items may be used to support the assertion that a test measures primarily a single construct. Substantial relationships of a test to other measures that are purportedly of the same construct and the weaknesses of relationships to measures that are purportedly of different constructs support the identification of constructs and the differences among them. Relationships among different methods of measurement and among various nontest variables similarly sharpen and elaborate the meaning and interpretation of constructs.
>
> Another line of evidence derives from analyses of individual responses. Questioning test takers about their performance strategies or responses to particular items or asking raters about their reasons for their ratings can yield hypotheses that enrich the definition of a construct. (AERA, APA & NCME, 1985, p. 10, emphasis added).

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

A test that is valid must be reliable. "Reliability refers to the degree to which test scores are free from measurement error" (AERA, APA & NCME, 1985, p. 19). As a basic textbook on testing points out, "The ceiling for possible validity of a test is set by its reliability" (Thorndike & Hagen , 1977, p. 87). In other words, if a test does not measure something reliably, it cannot be a valid measure of anything.

The reliability of the Massachusetts Teacher Tests

Absent sufficient data to assess the concurrent validity of the MTT, we decided to inquire into their reliability. Specifically, once we realized that relevant data might be available to us, we sought to examine the reliability of the scores on the April and July administrations. Comparing scores on these two administrations of the MTT might be thought of as a study of either test-retest or alternate-forms reliability. Test- retest reliability refers to the consistency of scores on two administrations of a test. Alternate-forms reliability refers to the consistency of scores on two different forms or versions of the same test. As we do not know to what extent the July MTT tests used identical questions as the April MTT tests, it is unclear whether our study should be termed a test-retest or alternate-forms study. But the essential idea is quite simple. It is to compare the scores on the MTT tests for people who took them in both April and July.

Adults' scores on basic skills tests of reading and writing tests should not change much over three months, and since no study guides were available, examinees could hardly have crammed for the second administration. So if the MTT tests are reasonably reliable, we would expect individuals' scores on these two administrations to be similar; if they are unreliable, we would expect the scores to vary widely.

We should acknowledge that there are several other ways in which test reliability might be estimated, such as internal consistency (indicating how much item results on one test administration tend to cohere.) (Note 5) But, the goal of certification tests such as the MTT is to estimate not simply examinees' competence on one test-taking occasion, but their competence in general. Alternate-forms reliability is thus more appropriate for assessing reliability. As Thorndike and Hagen (1977, p. 79) point out, "evidence based on equivalent test forms should usually be given the most weight in evaluating the reliability of a test."

After the July administration of the MTT, the Department of Education distributed lists of results for individual test-takers to institutions of higher education in the Commonwealth. When we realized that these lists contained individuals' scores for both the April and July tests, we decided to try to gather enough data to undertake a test-retest or alternate- forms reliability study. We contacted some institutions individually via phone, and some through the Commonwealth Education Deans' Council. Three institutions with large numbers of students retaking the MTT tests in July were contacted via letter, which read in part:

> We are seeking your cooperation in affording us access to data that will allow us to analyze some of the psychometric properties of the Massachusetts Teacher Tests.
>
> In particular we seek access to scores of examinees who took Massachusetts Teacher Tests in April and again in July. Access to these data will allow us to analyze the test-retest reliability of the Massachusetts Teacher Tests. Your institution received a computer printout labeled "Institution Roster Report By Test: Verified Institutional Affiliation" for test date July 11, 1998. We would like to receive either a copy of this computer printout or a data file containing relevant data. To preserve the confidentially of examinees we seek these data with the names and SSN's of candidates removed. (Letter to institutional representatives, September 22, 1998).

As of mid-December we had received MTT data from eight institutions, namely Boston College, Bridgewater State College, Elms College, Framingham State College, Lesley College, Salem State College, UMass Boston and Westfield State College. Five of these institutions are public and three are private. In both April and July, students from over 50 different institutions took the MTT. Eight out of

50 represents only 16% of the institutions that had students taking the MTT, but since these eight represent some of the largest teacher training institutions in the Commonwealth, they account for close to one third of the candidates who took the MTT in April.

Altogether we collected data on 219 people who took the MTT tests in both April and July, though not all 219 took the reading, writing and subject matter portions of the MTT on both occasions. (Note 6) One of the first things we noted about the April and July MTT scores is that some of the score changes seemed truly bizarre. (Note 7) For example, one individual was reported to have scored 36 on the reading test in April and 75 on the April writing test, but to have scored 89 on the reading test in July and 17 on the writing test. In another case, an individual was reported to have scored 56 on the writing test in April and then got an 11 in July. Such huge score changes seemed so unlikely that we inquired into the accuracy of the reported scores. In both of these cases, the scores reported were verified by institutional representatives. In the first case we were told that the individual had not been prepared for the reading test in April, and that the dramatic increase from 36 in April to 89 in July was explained by the fact that the test taker had known that the latter counted toward certification. Why did the writing score plummet from 75 to 17 while the reading score increased from 36 to 89? According to the institutional representative, this happened because the test-taker started taking the July writing test, but then remembered that because she had scored more than 70 in April, she did not have to take the writing test again in July. Hence she simply stopped answering questions. Nonetheless the July score of 17 was reported to the institution as a failure. In the second case, in which writing scores dropped from 56 in April to 11 in July, the institutional representative verified the accuracy of the scores. She had no explanation for the dramatic score decrease, but added that the individual who had received these scores had left Massachusetts to take a teaching job in Arizona.

Table 1 presents the summary statistics for the 219 cases of April and July MTT test-takers for which we have data.

Table 1: Summary Descriptive Statistics on April-July MTT Test-Takers

	Reading 4/98	Writing 4/98	Reading 7/98	Writing 7/98
Count	215	218	130	173
Mean	65.2	63.1	69.4	70.7
Median	66	63.5	70	71
Standard deviation	14.7	10.3	15.2	11.8
Minimum	3	36	21	11
Maximum	93	87	96	96

These data suggest that this sample is not unlike the April MTT test takers in general. On average, they fell below the passing score of 70 on both the reading and writing tests. Initially, these results would appear to make the MTT results seem reasonably reliable. Among repeat test-takers, the average reading scores increased from 65.2 to 69.4, and the average writing test scores from 63.1 to 70.7, apparently modest changes. But note the differences in the count of people in this sample who took the April and July tests. While more than 200 took both reading and writing tests in April, fewer than 180 took the tests in July. This reflects the fact that test-takers had to retake tests in July only if they had scored less than 70 on either the reading or writing tests.

Hence, in order to assess the reliability of the MTT tests, we need to examine the correlations between April and July tests for examinees who took the same portions of the MTT tests on both occasions. Table 2 shows the intercorrelations of reading and writing test scores for people who took both tests. For those who took the reading test in April and July, the correlation of scores was 0.29; for writing, 0.37.

Table 2: Intercorrelations of April and July MTT Scores

	Reading 4/98	Writing 4/98	Reading 7/98	Writing 7/98
Reading 4/98	1.00 (215)	0.07 (215)	**0.29** **(127)**	0.24 (169)
Writing 4/98		1.00 (218)	0.47 (129)	**0.37** **(172)**
Reading 7/98			1.00 (130)	0.06 (94)
Writing 7/98				1.00 (173)

Note: Sample sizes shown in parentheses, test-retest correlations in bold.

These test-retest intercorrelations are extraordinarily low. Correlation coefficients can range from -1.00 to +1.00. Test-retest correlation coefficients for well-developed standardized tests typically range between 0.80 and 0.90. For example, test-retest correlations for the SAT have been reported to range between 0.86 and 0.90 (Donlon, 1984, p. 54). Similarly, Thorndike & Hagen (1977, p. 321) report alternate-form reliability coefficients in the range of .85 to .95 for the Stanford Binet. In contrast, the scores of examinees who took MTT reading tests in April and July correlated only 0.29, those of examinees who took the writing test in April and July correlated 0.37.

To provide a more detailed picture of the relationship between April and July scores, Figures 1 and 2 show scatterplots of test scores for individuals in our sample who took the reading and of writing tests, respectively, on both occasions. Several patterns are apparent in comparing these figures. First, note that the "scatter" in reading test scores is greater than that in writing scores. This simply reflects the findings shown in Table 2 above; namely, that the correlation between reading scores in April and July (0.29) was smaller than that for writing test scores (0.37).

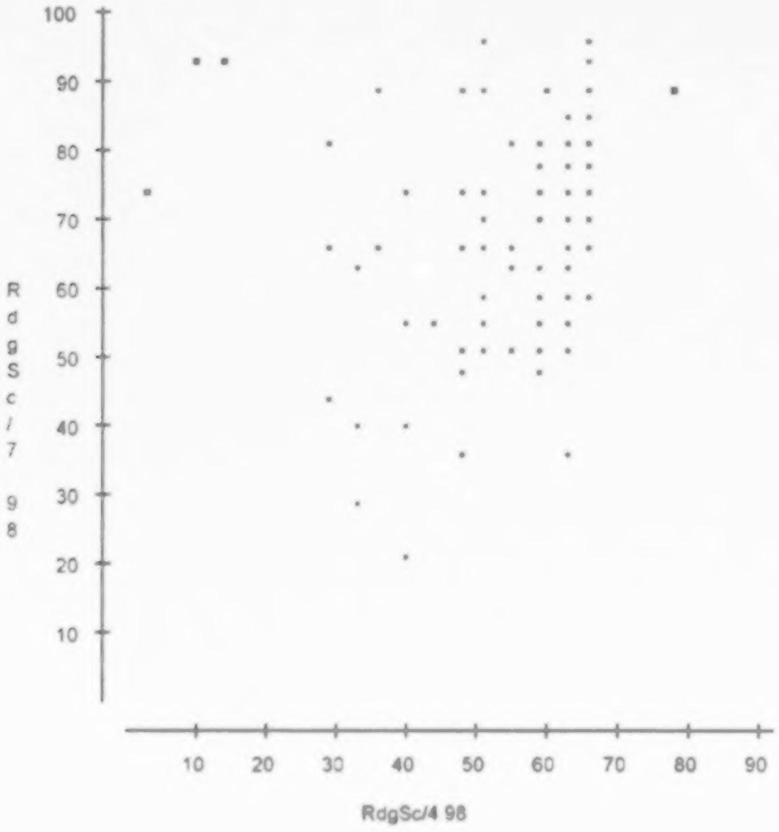


Figure 1. Scatterplot of April and July MTT Reading Test Scores

Note also how widely retest scores vary among people who had approximately the same test scores in April. For example, in Figure 1, among examinees who had scores of about 60 on the reading test in April, retest scores in July ranged from less than 40 to about 90. And, as is apparent in Figure 2, among test-takers who scored in the 65 to 69 range in April, retest scores range from about 50 to 90.

These figures also illustrate some of the huge score changes that initially caught our attention. These cases, often called "outliers" in data analysis, are marked with x's in Figures 1 and 2. In Figure 1, for example, note the three cases in the upper left corner. In all three cases, examinees had scores of less than 20 on the reading test in April but more than 70 in July, increases of more than 3 standard deviations. And in Figure 2, note the case in the lower right corner, representing someone who had a score of 75 in April, but a score of 17 in July. This is the case mentioned previously that was so bizarre that we asked the institutional representative to verify the accuracy of the data--the case of the test-taker who, remembering she did not have to take the writing test again, simply stopped answering questions . (Note 8) The other clear "outlier" in Figure 2 is lowest x on the figure, representing someone who had a score of 56 on the writing test in April, but 11 in July.

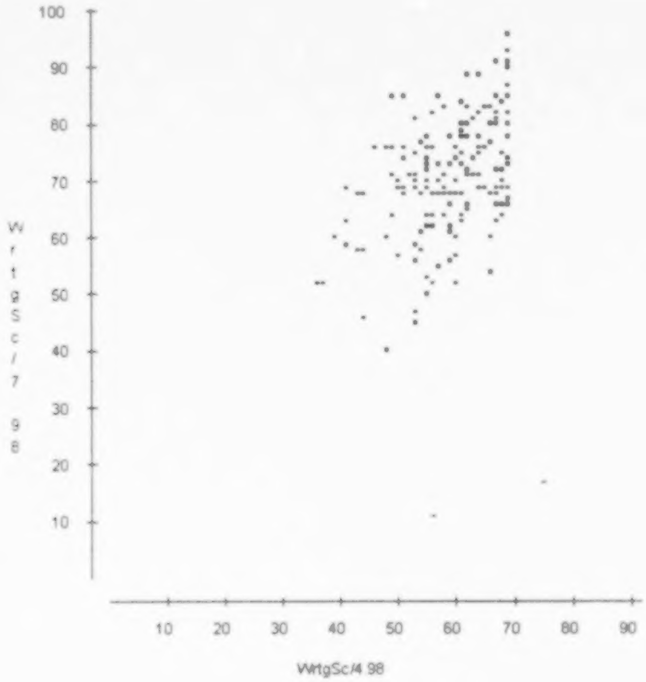


Figure 2. Scatterplot of April and July MTT Writing Test Scores

We have checked these "outlier" cases and all are accurate in terms of scores reported to institutions. Nonetheless, as a more conservative examination of the test-retest reliability, we recalculated the test-retest correlations after deleting the outliers. We refer to these groups, after deleting outliers, as our trimmed samples. Specifically, after deleting the four unusual cases marked in Figure 1 with x's, the test-retest correlation for the reading test rose to 0.49. Similarly, after deleting the two outlying cases shown in Figure 2, the test-retest correlation for the writing test increased to 0.48.

This brings us to one other feature apparent in Figures 1 and 2, and also to a possible explanation for the remarkably low test-retest correlations shown in Table 1. Note that in both Figures 1 and 2, there is only one case for which retest data are available for an examinee who scored 70 or above in April. This is because people who scored 70 or above "passed" the tests and did not have to retake them in order to be provisionally certified. With this one exception, our test-retest data for the MTT are for people who scored below 70 on the April tests. This leads to one possible explanation for the unusually low test-retest correlations, namely attenuation of observed correlation coefficients due to restriction of range. This concept is easy to explain with an example. People's height tends to be correlated with their weight. Tall people tend to weigh more than short people. Thus, we would find a positive correlation between the heights and weights of adults in general. But suppose that we consider a sample of people who are all exactly five feet tall. If we examine the correlation between their height and weight, we will find a zero correlation for the simple reason that they are all of the same height. By focusing on people who are exactly five feet tall, we have restricted the range on this variable; hence, the observed correlation between height and weight for this sample has been reduced or attenuated. This is what is meant by attenuation due to restriction of range. If we restrict the range of a variable, the observed correlation between this variable and another will be attenuated, as compared to the correlation that likely would be observed if the range on the variable were not restricted.

Hence, before concluding that the MTT reading and writing tests are unreliable, we need to consider the possibility that attenuation due to restriction of range, with most of test-retest data available only for examinees who scored less than 70 on the April tests, may have led to the low test-retest correlations shown in Table 2. Fortunately, the phenomenon of attenuation of correlation coefficients due to restriction of range has been widely recognized in the testing and measurement literature. Formulas and tables are available to allow estimation of "unattenuated" correlation coefficients when restriction of range is taken into account (slightly different formulas are available, for example, in Lord & Novick, 1968; Cronbach, 1971; and Linn, 1982).

Lord & Novick (1968) present an extended discussion of attenuation due to restriction of range and tables showing how observed correlations can be corrected for attenuation. If we assume that the relationship between two variables is linear and that the conditional variance of one does not depend on the particular value of the other (the assumption of homoscedasticity), then the following table shows the corrections for observed correlations when the percentage of the sample is restricted to the top (or bottom) 60%, 50%, 40% and 30% of the entire population. As shown in Table 2 above, we found that the observed correlation between the April and July MTT reading tests was 0.29. However, 70% of examinees passed the April reading test, so the range of examinees who had to take the July reading test was "restricted" to only the bottom 30% of the population of April examinees. Table 3 indicates that a correlation of 0.30 observed when range is restricted to 30% of a population would be corrected to 0.519 for the whole population. Similarly, we observed a correlation of 0.37 between scores on the April and July writing test, but since about 60% of examinees passed the April writing test, the group retaking the July writing test was restricted to about 40% of the population. Table 3 indicates that an observed correlation of 0.40 in a sample restricted to 40% of a population would be corrected to 0.616 for the entire population. For the trimmed samples, the observed correlations of 0.49 and 0.48, would be corrected to about 0.74 and 0.72, again presuming that only the bottom 30% retook the reading test and the bottom 40% retook the writing test.

Table 3: Corrections for Attenuation Due to Restriction of Range

Normal deviate z	Percent selected in restricted sample	Standard devia-tion in selected group	Ratio of SD in unselected to SD in selected groups (K)	Observed correlation of 0.30 in restricted sample corrected to	Observed correlation of 0.40 in restricted sample corrected to	Observed correlation of 0.50 in restricted sample corrected to
-0.25	59.9	0.65	1.54	0.436	0.558	0.644
0	50	0.6	1.64	0.458	0.582	0.688
0.25	40.1	0.56	1.79	0.491	0.616	0.719
0.5	30.8	0.52	1.93	0.519	0.644	0.744

Source: Adapted from Lord & Novick, 1968, pp. 140-142.

To verify these corrections for attenuation due to restriction of range, we conducted simulation analyses to address questions such as the following. If the test-retest correlation among a group of test takers was 0.50, what would be the correlation observed if only the bottom 30% on the initial test were considered? We do not attempt to present all of the results of these simulations here, but instead, in Figure 3, present the results of one iteration of the data simulations aimed at addressing the following question. If there were a test- retest correlation between test 1 (t1) and re-test (t2) of 0.50, what would be the observed correlation between test and re-test scores if attention were restricted to only the bottom 30% on the initial test (t1). What our results show is that if there were a test-retest correlation of 0.50 among the entire population of test-takers, restricting attention to only the bottom 30% of test-takers on the initial test (t1) would reduce (or attenuate) the observed correlation to about 0.30. These results confirm the theoretical results reported above. Given that we observed a test-retest correlation of about 0.30 in the 30-40% of examinees who had to retake the MTT, our estimate of the test-retest correlation for the MTT, if all examinees had retaken the tests, is about 0.50.

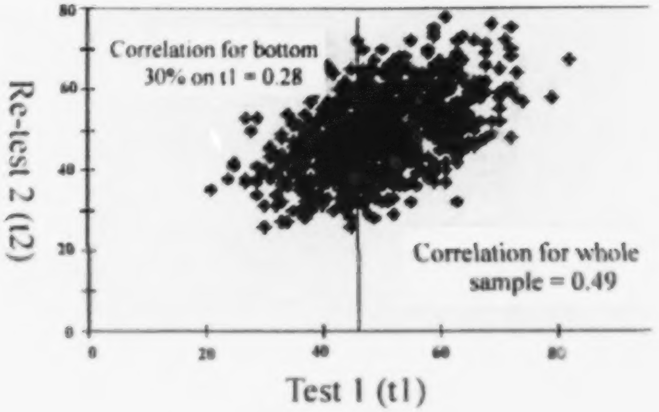


Figure 3. Example of Simulation Results

Note: Results shown here are for a sample of 1000

Test-retest correlations in the range of 0.50 (or even 0.70) are unusually low. In comparison, as previously mentioned, test-retest correlations for the SAT have routinely been found to be in the range of 0.85 to 0.90 (Donlon, 1984, p. 54). There are several ways of illustrating the implications of test-retest reliability being as low as 0.50. One way of interpreting a test-retest reliability coefficient rtt is as the ratio of signal to "noise plus signal," or as the ratio of true score variance to observed score variance.

$\mathbf{r}_{tt}$ = signal / (signal + noise) = true score variance / observed score variance

Since observed score variance is composed of true score variance plus error score variance (see Anastasi, 1976, pp. 120-22, or many other textbooks on testing, for more detailed explanations), this equation can also be expressed as
$\mathbf{r}_{tt}$ = (true score variance) / (true score variance + error score variance)

Thus, it is easy to see that when a test-retest reliability coefficient $\mathbf{r}_{tt}$ is as low as 0.50, observed scores are composed of as much error score variance as of true score variance. Thus a test-retest correlation of 0.50 indicates that MTT scores contain as much error as true score variance. Even a test-retest correlation of 0.70 indicates that MTT scores are composed of 30 percent error variance.

A second way of showing the meaning of a test-retest reliability coefficient $\mathbf{r}_{tt}$ is to use it to calculate the standard error of measurement, as follows:

$$e = s_t \sqrt{1 - r_{tt}}$$

(Thorndike & Hagen, 1977, p. 85; Anastasi, 1976, p. 128)

where:

e = standard error of measurement

s_t = standard deviation of test scores, and

r_{tt} = test-retest reliability coefficient.

As shown in Table 2, in our test-retest sample, we found the standard deviations of reading and writing test scores to be about 15 and 11 points respectively. However, these observed standard deviations were based on the restricted sample of retest examinees (with only 30% of April examinees having to retake the reading test and 40% the writing test), so we need to find a way of estimating the standard deviations of MTT test scores for the entire population of test takers.

As we have pointed out, even after the MTT have been administered four times, over a period of a year, no technical report on these new tests has been issued. Hence, we must rely on data shared with us by cooperating institutions to estimate the standard deviations of MTT reading and writing test scores among the entire population of examinees. We have available two different avenues for pursuing this end; usingtheoretical adjustments of data on our test-retest sample and using data institutions shared with us on all their students who took the MTT in April and July.

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

Using the theoretical approach (and the data shown in Table 3 above), we can multiply the restricted sample standard deviations by 1.93 and 1.79 to estimate the standard deviations in the full population of April examinees. Since 15 x 1.93 = 28.95, and 11 x 1.79 = 19.69, we may use these figures as one set of estimates of the standard deviations of the MTT reading and writing tests. A second approach is to examine the standard deviations of the April tests for the institutions which gave us data on all of their April test-takers. We found that that the within-institution standard deviations to be as high as 19 points for the April reading test and 16 for the April writing test.

Hence, as summary estimates of the standard deviations of the April tests, we averaged these two estimates, which yielded 24 [(29 +19)/2] and 18 [(16 +20)/2] as ballpark estimates of the standard deviations of the April tests for the full population of test takers. Then we estimate standard error of measurement for the MTT reading and writing tests as follows:

$$e_r = 24\sqrt{1 - 0.519} = 24\sqrt{0.481} = 24 \times 0.694 = 16.6$$

$$e_w = 18\sqrt{1 - 0.616} = 18\sqrt{0.384} = 18 \times 0.62 = 11.2$$

Even if we use the more conservative estimations of test-retest correlations, based on the trimmed samples (that is, with outliers deleted) and adjusting for attenuation dues to restriction of range, namely 0.74 and 0.72 for the MTT reading and writing tests respectively, these would still imply standard error of measurement of 12.2 and 9.5. In other words, our results suggest that the standard errors of measurement in the April MTT Reading and Writing tests were about 17 and 11 points respectively (or at best 12 and 9). While neither the Massachusetts DOE nor NES has yet released any technical information on the scaling of the

MTT, we have found MTT scores to range from near zero to almost 100. If indeed the scores for the MTT are on a 100 point scale, this means that standard errors of measurement of 9 and 17 points represent some 9% to 17% of the entire score range. This means that examinees scoring in the range of 50 to 69 may easily have "failed" the MTT simply because of measurement error, and, conversely, ones scoring in the range of 70 to 90 may well have "passed" simply because of the large degree of measurement error in the MTT tests.

These errors of measurement on the MTT may be compared with the standard error of measurement on well-known tests for which technical documentation is available. The SAT (originally, the Scholastic Aptitude Test, briefly renamed the Scholastic Assessment Test, and now just the SAT) is reported on a scale that ranges from 200 to 800, or 600 points. The standard errors of measurement of the SAT verbal and quantitative scores have been reported to be in the range of 29-34 points (Donlon, 1984, pp. 33-34), or 4.4 to 5.7% of the score range. The standard errors of measurement for the Graduate Record Examination have been reported to be 33, 38 and 36 points for the GRE Verbal, Quantitative and Analytical subtests respectively (Conrad, Trismen & Miller, 1977, p. 19). Since these scores are reported on scales ranging from 600 to 670 points, these standard errors of measurement are all less than 6% of scale range. If the standard errors of measurement of MTT scores are 9 to 17 points, as we have estimated, representing 9% and 17% of the MTT score range, this means that MTT scores have almost double to triple the degree of error as the SAT and the GRE (as estimated by SEM relative to scaled score range).

The reliability of classifications based on the MTT

This brings us to a point mentioned in the introduction of this report. As the 1985 *Standards for Educational and Psychological Testing* point out, for licensure or certification tests on which people are rated as passing or failing, it is important to provide data not just on test scores, but also on the reliability of classification decisions based on those scores (Standard 11.3, p. 65). As Ron Hambleton has pointed out, "there were serious problems with the setting of passing scores on the reading literacy, writing and subject matter [MTT] tests":

1. A detailed description of what it means to be qualified was not developed;
2. Panelists who set passing scores did not have an opportunity to discuss their recommendations with each other prior to finalizing their recommendations to the Board;
3. Technical data arising from the process of setting passing scores was not presented to the Board for their consideration. (Hambleton, 1999, pp. 20-21)

Nonetheless, data available from the April and July administrations of the MTT allow us to examine the reliability of pass/fail classifications based on the MTT reading and writing tests. In a report entitled "Massachusetts Teacher Tests. Summary of Institution results for Second-Time Test Takers. Test Date: July 11, 1998. Test Summary," the Massachusetts DOE released data shedding direct evidence on this point. The report lists the number, and percent passing, of examinees who retook the MTT on July 11, 1998. Results are reported only for institutions that had more than four candidates. Hence this table showed reading test results for only 18 institutions and writing test results for 23 institutions. These data are shown in Table 4 below.

Table 4: Passing Rates of Second-Time Test Takers on July 11, 1998, MTT Tests

Institution	Reading N	Reading % pass	Writing N	Writing % pass
American International College	10	70.0	13	30.8
Anna Maria College	5	20.0	5	40.0
Assumption College		6	50.0	
Boston College	7	85.7	5	80.0
Bridgewater State College	41	58.5	59	55.2
Curry College	6	33.3	6	16.7
Fitchburg State College	25	56.0	31	45.7
Framingham State College	16	31.3	19	52.6
Lesley College	12	66.7	14	64.3
Mass. College of Liberal Arts	6	33.3	10	50.0
Merrimack College	5	40.0	9	66.7
Salem State College	17	76.5	26	76.9
Simmons College		6	100.0	
Springfield College	22	36.4	27	48.1
Stonehill College	10	90.0	13	53.8
Univ. of Massachusetts/ Amherst	16	68.8	20	35.0
Univ. of Massachusetts/ Boston		6	50.0	
Univ. of Massachusetts/ Dartmouth		6	50.0	
Univ. of Massachusetts/ Lowell	5	60.0	13	76.9
Westfield State College	27	48.1	32	34.4
Wheelock College	6	50.0	12	50.0
Worcester State College	11	72.7	15	60.0
Unaffiliated candidates	35	60.0	47	44.7
Mean		55.6		53.6
Median		58.5		50.0

Source: Adapted from Mass DOE, "Massachusetts Teacher Tests. Summary of Institution results for Second-Time Test Takers. Test Date: July 11, 1998. Test Summary."; (Available at http://www.doe.mass.edu/teachertest/)

As the data in Table 4 indicate, the mean pass rate (unweighted average across institutions for which data were reported) among second-time MTT test-takers was over 50% on both the reading and writing tests. Though we do not show weighted results in Table 4, these data indicate that 160 of 282 or 57% of examinees taking the reading test for the second time passed, and 207 of 400 or 52% of those taking the writing test passed. This indicates that the misclassification rate among those who "failed" the April tests was over 50% on both the reading test and the writing test. This seems extraordinarily high given that adults' basic skills in reading and writing are unlikely to change much over a three month-period (and as previously mentioned, candidates could not cram for the July test). Note, too, that the misclassification rate was higher on the reading test than on the writing test-- exactly what would be predicted from the results of our reliability analysis, which showed the reading test to be less reliable than the writing test.

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

What these results do not show, of course, is the rate of misclassification of those who passed the MTT tests. Nonetheless, it seems certain, given the results of our analyses, that a substantial proportion of those who "passed" the MTT reading and writing tests by 10 to 20 points did so simply because of test unreliability. People who scored above 69 on the MTT reading and writing tests, and thus "passed" these tests in April did not have to retake them in July. Hence we have no direct way of estimating the false "pass" rate. But to get a rough idea of this kind of misclassification we examined the percentage of April examinees scoring in the 50-69 point range whose scores decreased. We found that some 4% to 10% of first-time test-takers in these score ranges had decreased scores upon retest. Thus, a very conservative estimate of the percentage of April examinees who "passed" simply because of measurement error would be 2% to 5%.

Table 5: Passing Rates of Second-Time Test Takers Reported by DOE compared with Retest Sample

	Report	ed by	Mass	DOE	Ad	Hoc	Retest	Sample
Institution	Rdg N	Rdg % pass	Wrtg N	Wrtg % pass	Rdg N	Rdg % pass	Wrtg N	Wrtg % pass
Boston College	7	85.7	5	80.0	6	83.3	5	80.0
Bridgewater State College	41	58.5	59	55.2	42	59.5	60	56.7
Framingham State College	16	31.3	19	52.6	16	31.3	19	52.6
Lesley College	12	66.7	14	64.3	12	66.7	14	64.3
Salem State College	17	76.5	26	76.9	21	66.7	34	79.4
U. Mass/Boston			6	50.0			6	50.0
Westfield State College	27	48.1	32	34.4	27	48.1	32	34.4

Source: Adapted from Mass DOE, "Massachusetts Teacher Tests. Summary of Institution results for Second-Time Test Takers. Test Date: July 11, 1998. Test Summary". (Available at http://www.doe.mass.edu/teachertest/)

The data summarized in Table 4 also allowed us to check the findings from our test- retest sample against passing rates reported by the DOE that are summarized in Table 5. This table presents the passing rates reported by the DOE with those apparent in our test- retest sample. Note first that this table shows no results for Elms College; the DOE did not report any results for this institution because it had fewer than five second-time test takers. For four of the remaining seven institutions, the sample sizes (Ns) and passing rates reported by the DOE are exactly the same as those in our test-retest sample. For the remaining three institutions there are slight differences between results reported by the DOE and those apparent in our test-retest sample. For Boston College, the DOE reported seven second-time takers for the reading test, whereas we counted only six in our test-retest sample. We have examined the data for Boston College in detail and suspect that this discrepancy arises from an unusual case in which one student took the MTT in April, and had an April writing test score reported, but had no April reading test score reported in results transmitted to Boston College. Thus, apparently this individual was counted in the DOE results as a second-time test-taker, but was not included in our test-retest sample because no reading test score was reported for April. The other two institutions for which there are slight discrepancies are Bridgewater and Salem. In both cases, the Ns for the test-retest sample are slightly higher than the Ns reported in DOE results. These differences apparently derive from the fact that the DOE results are reported only for individuals whose institutional affiliation was verified by the institution by a particular date. The reason for the slightly larger Ns for Bridgewater and Salem is that the data provided to us apparently included a small number of cases that were treated as unaffiliated examinees by the DOE.

Indeed, the DOE's policy of institutional affiliation of MTT test takers seems to be of doubtful merit and of changing meaning. In reporting institutional results for the April, July, and October administrations of the MTT, the DOE has offered a number of slightly different "Interpretive cautions and notes." But in each instance, the first has read as follows:

> 1. Information regarding candidate institutional affiliation was obtained from candidates as self-reported information on the registration form during the test registration process. This information was forwarded to institutions of higher education, which were provided with an opportunity to verify the candidates' institutional affiliation. The institutions were informed that if they did not respond to the verification request as explained, the data to be included in their results would be based on candidate-reported institutional affiliation.

The institutional results for the April administration of the MTT were released under a memo from David Driscoll dated July 21, 1998. Together with institutional results for the July and October 1998 administrations, they are also available on the DOE web site. When we examined results for April and July, we noticed that there was a sharp increase in the number of "unaffiliated" candidates, that is, ones whose affiliation was with institutions outside Massachusetts or was not verified by the institutions concerned. Hence, we used the data available from the DOE web site to calculate the percentages of test-takers at each administration who were listed as "unaffiliated." As the results in Table 6 show, between April and July there was a fourfold increase in the percentage of test-takers listed as unaffiliated.

Table 6: First-time Test-Takers Listed as Unaffiliated

MTT Test	APRIL	JULY	OCTOBER
Reading	227 / 1794= 12.7%	891 / 1702= 52.4%	778 / 1533= 50.8%
Writing	227 / 1808= 12.5%	898 / 1707= 52.6%	783 / 1544= 50.7%

Sources:
http://www.doe.mass.edu/teachertest/7981st.html
http://www.doe.mass.edu/teachertest/summary498.html_
http://www.doe.mass.edu/teachertest/1098inst/1test.html
(Data summarized in Table 6 were downloaded 1/5/99)

One possible explanation for this sharp increase is that students enrolled in out-of- state colleges were more readily able to come to Massachusetts to take the MTT in July than they were in April. This is not confirmed, however, by the fact the proportion of first-time test-takers listed as unaffiliated remained very high, more than 50%, in the October administration. Thus, what appears to have happened is that Massachusetts institutions of higher education with teacher preparation programs changed the manner in which they verified the affiliation of students after the first administration of the MTT. Indeed, on September 26, 1998, an article in the Boston Globe, "BU test to screen teacher hopefuls, disclaim failures," reported that Boston University had instituted a policy of not verifying students' affiliation with BU unless they had passed a literacy screening test before taking the MTT (Zernike, 1998).

What the BU policy and the fourfold increase in the percentages of unaffiliated candidates indicate, however, is that even if the MTT test scores were reliable and valid, the DOE's practice of publishing "institutional" results may be highly misleading unless some better and uniform methods of "verifying" candidates' affiliation with institutions is developed. And even if that problem were solved, the ranking of schools based on student test scores is of doubtful merit.

Ranking schools and school districts (and even states and countries) on student test results seems to be increasingly popular with the media in recent years. This is both unfair and ineffective in improving education. It is unfair for the simple reason that judging the effectiveness of educational institutions should be based not on end-of-school test scores, but on "value-added" as a result of experience in the school. If a value-added perspective is not adopted, then highly selective institutions (such as Harvard, Boston College, and Boston University among teacher preparation institutions in Massachusetts) may come out looking good in such rankings, simply because they only admit students who are good test-takers to begin with, not because of how much students learn while attending them.

And ranking of schools based on student test scores can be both unfair and ineffective unless attention is paid to not just test scores as "outcomes," but also to the educational processes that produced those outcomes. In recent years, for example, numerous cases have been revealed in which schools cheated on tests in order to make their rankings look better. And even absent such manipulation of test scores, lack of attention to processes provides little leverage for improvement (Haney & Raczek, 1993).

The content validity of the Massachusetts Teacher Tests

Validity refers to the appropriateness and meaningfulness of inferences drawn from test scores. As explained previously, three types of evidence have been recognized in this connection; content-, criterion- and construct-related validity evidence. The Ad Hoc Committee originally intended to gather one form of criterion-related evidence, namely concurrent validity evidence that would compare scores on the MTT with those on well-established tests for college graduates. Having failed to gather enough comparable scores to allow reasonable statistical analysis, we undertook the reliability studies described above. Our findings that the MTT reading and writing tests are unreliable and have led to remarkably high rates of misclassification obviously cast doubt on their validity, but it is useful also to comment directly on the content and construct validity of the MTT tests.

In general, content validity refers to whether test questions cover the right material, that is some defined domain of content. For licensure and certification tests, this translates into whether test questions clearly and correctly span the domain of knowledge necessary to protect the public from people who are not competent. Since competence in most professional fields is hard to define and measure precisely, content validation studies of licensure and certification tests are usually based on the expert judgment of people in the field being tested. In the case of the MTT the relevant fields were those of education and teaching. Typically in content validation studies, test developers ask practitioners to judge whether test questions are job-related and whether they match a particular content domain (often defined in terms of test objectives).

Several Massachusetts officials have said publicly that such content validation studies have been done by panels of educators across the state in the development of the MTT. However, no relevant reports or documentation on these reviews have yet been released, even though the MTT have now been administered four times.

At the same time, a particular portion of the MTT gives us pause about the way in which the content validation and job-relatedness studies have been used in the development of the new MTT. On the first MTT, administered in April 1998, as part of the writing test, examinees were asked to transcribe a 156-word text drawn from the Federalist papers (written by James Madison in 1787) as the text was read three times by a narrator on audiotape. (According to an August 5 story in the Boston Globe, the dictation exercise was suggested by Massachusetts Board of Education members Edward Delattre and John Silber; Hart, 8/5/98).

It seems to us highly implausible that such an exercise would be judged a valid and job-related measure of writing competence by a majority of panelists reviewing content validity. Also, though we have reviewed more than 50 years of teacher competency testing in the United States (the NTE, for example was created in 1940; Haney, Madaus and Kreitzer, 1987), we have found no other instance in which a dictation exercise has been used as a measure of teacher competence.

How then could such an unusual exercise have shown up on the MTT? We cannot be sure. But it is worth noting that in the Alabama teacher testing case referred to previously (which is reproduced in part in appendix 2), Judge Myron Thompson found that on the Alabama Initial Certification Test "a significant number of items appearing on the examinations failed to reflect accurately the collective judgment of curriculum committee members. In some cases changes to actual test items were not implemented. In other cases, items that had never been reviewed by a curriculum committee appeared on examinations. . . . [Also,] many items appeared on the examinations even after they had been rated content invalid by the requisite number of Alabama panelists" (Richardson v. Lamar County Bd. of Educ. 729 F. Supp 806, 821-822). It may be recalled that the developer of the Alabama test is the same company that developed the MTT.

The construct validity of the Massachusetts Teacher Tests

A third and more general form of validity evidence of the meaning of test scores relates to the "constructs" that the scores represent. As the 1985 *Standards* point out, "Substantial relationships of a test to other measures that are purportedly of the same construct and the weaknesses of relationships to measures that are purportedly of different constructs support the identification of constructs and the differences among them" (AERA, APA & NCME, 1985, p. 10).

In carrying out test-retest analyses, we were surprised to find that in our sample of people who took the MTT in both April and July, there was a correlation of less than 0.10 between MTT reading and writing test scores in both April and July (see Table 2). Since reading and writing are both verbal or literacy skills, we would expect to find substantial correlations between test scores of these related constructs. But, as we have noted, the test-retest sample was restricted (with one odd exception) to people who had failed either the reading or writing test in April. Thus the group of repeat test- takers represents a highly restricted or attenuated sample of MTT test-takers in general.

To examine the relationship between MTT reading and writing scores on less restricted groups of test-takers, we returned to data obtained from for analyzing test-retest reliability. Several institutions had provided us with data on all of their students who had taken the MTT in April and in July. These data allowed us to examine the MTT Reading x MTT Writing correlations on larger samples than when we confined ourselves to individuals who took the MTT in both April and July. Table 7 presents results of these analyses.

Table 7: Correlations of MTT Reading and Writing Scores

	April	July
Boston College	0.42 (111)	0.20 (44)
Lesley College	0.56 (62)	0.65 (60)
Westfield State	0.57 (101)	0.50 (107)
Total N	274	211
Median r	0.56	0.50

Note: Sample sizes shown in parentheses

Note, first, that the correlations between MTT Reading and MTT Writing scores vary somewhat: from 0.42 to 0.57 for April and from 0.20 to 0.65 for July. Part of this is due surely to sample size. For example, the most anomalous correlation in Table 7 is for Boston College for July test results (correlation of 0.20). Note, however, that for this sample, there was an N of only 44. If we consider only those cases in which N>100, we see a much more consistent pattern, with MTT Reading x MTT Writing correlations of 0.42, 0.50, and 0.57. This suggests that the average correlation between MTT Reading and MTT Writing test scores is about 0.50.

This finding may be compared with previous research on the intercorrelations between measures of two verbal skills. Cronbach (1970), for example, reports that the Verbal and Spelling subtest scores on the General Aptitude Test Battery (GATB) correlate in the range of 0.66 to 0.72. Donlon reports that Test of Standard Written English (TSWE) scores correlate with SAT Verbal scores in the range of 0.76 to 0.80 and with SAT Reading scores in the range of 0.72 to 0.77 (Donlon, 1984, p. 81). Similarly, Conrad, Trismen and Miller report that GRE Verbal and

~ EPAA Vol. 7 No. 4: Haney, Fowler, Wheelock, Bebell & Malec "Massachusetts Teacher Test" ~

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

Interviews with MTT-Takers: Vignettes and Summary

Several individuals who sent copies of their MTT score reports to the Ad Hoc Committee spontaneously offered us comments on the new MTT. For example, one woman wrote:

> After graduation from xx College--one of the best schools in the area, ALL of my daughter's friends failed at least one section of the MA test. Is something wrong with this picture? After a phone chain among many parents, we all agree there is a problem with the MA test, as these students did all pass required testing for other states. [Underline in original; name of College deleted to protect confidentiality.]

Another correspondent wrote:

> My scores on the Praxis series earned me a license to teach Language Arts and Social studies to grades six through nine in North Carolina. Unfortunately, this was not enough to earn a reprieve from the Massachusetts test. This is just one of the aspects of the test with which I take issue.
>
> One major problem with the Massachusetts Teacher Tests is that there is no preparation offered. When I called to request information on the test, I received a packet of test objectives for each of the tests I was taking. This information was practically useless, as I still had no clue as to the format of the test. [Letter to the Ad Hoc Committee, dated August 20, 1998.]

In light of such comments, we contacted the first 15 test-takers who had sent us copies of their score reports, to ask whether they were be willing to be interviewed, on condition that we keep their identities confidential. All 15 agreed. The interviews focused on the current professional status of the teacher candidates, sought their views on the administration and content of the MTT tests and asked about their attitudes toward testing and teaching. We gathered this information in telephone interviews lasting between one half to one hour during November 1998. We took notes during phone conversations and elaborated them after the end of conversations. Typewritten accounts of each interview were then prepared, and results across interviewees were analyzed by looking for common themes and comments.

Interview sample

Although this was a small, self-selected sample, those who agreed to an interview represented a wide range of experiences. Of the fifteen, seven (47%)

passed all three parts of the test (reading, writing, and subject area) on their first try, approximating the passing rate for the state overall. Two additional candidates (13 %) passed both literacy sections, but failed their subject area tests. Five (33%) passed one portion of the literacy section only, with two of them also passing their subject area. Only one out of 15 candidates interviewed failed all three portions of the test.

The 15 candidates had college degrees from nine private and four public colleges and universities, with two unknown. Although most had received a first degree in 1998, several were teachers who had moved to Massachusetts after teaching in other states, and one had 20 years experience as a teacher.

At the time of the interviews, eight of the fifteen were certified to teach in Massachusetts; eleven were certified in at least one other state--including Connecticut, New York, Rhode Island, New Hampshire, New Jersey, California, Georgia, Arkansas, Missouri, North Carolina, Maryland, and Tennessee. To receive such certification, candidates typically had passed a required test. Nine of them had taken the National Teachers Examination (NTE), recently renamed and broadened to become the Praxis. Others had taken specific state tests, the Graduate Record Examinations, or the Millers Analogies Test (MAT).

Four took the MTT in April 1998, during the test's first administration. Nine took the MTT for the first time during the second round of testing, in July 1998. Two took it in October 1998. All candidates interviewed submitted scores for the morning two-part literacy portion of the test. Candidates also submitted scores for elementary education (6), English (2), physical education (2), and general science (1), physics (1), music (1), middle school (1), and special needs (1).

Although a common rationale for teacher certification tests such as the MTT is that they will protect schools, parents and school children from incompetent teachers, the MTT tests did not prevent most teacher candidates in the sample from securing work in some kind of teaching capacity. Regardless of whether they passed or failed the MTT, 12 of the 15 candidates interviewed currently work in public, private, parochial, and charter schools, both in and out of state.

Of the seven candidates who passed the MTT, two are working in full-time teaching positions in Massachusetts public schools, while two are working in-state as long-term public school substitutes. Two more work as full-time teachers in public schools out of state. One candidate is not working by choice.

Table 8: Current Employment Status of Interviewees

	PASSED MTT (n=7)			**FAILED MTT (n=8)**		
Employment Status	Full time Teacher	Long Term Substitute	Teacher's Aide	Full time Teacher	Long Term Substitute	Teacher's Aide
Employed in education						
In Mass.						
Public School	2	2				1
Private School				1		
Charter School				1		
Out of State						
Public School	2			1		
Private or parochial school				2		
Employed, not in education				1		
Not employed	1			1		

Of the eight candidates who "failed" the MTT, six are working in schools. One works full-time in a Massachusetts charter school; a second works as a teacher in a Massachusetts private school; and a third works in the Commonwealth as a full-time public school teacher's aide. Of the remaining four, three work as full- time teachers out of state, one each in a public, private, and parochial school. One candidate is working in the travel industry, and one is not working by choice.

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

Vignettes

Before summarizing the general findings from our 15 interviews, it is useful to provide vignettes of two candidates, to highlight the diversity of candidates and their experiences. Pseudonyms are used in these vignettes to maintain our agreement of confidentiality with interviewees. (Note 9)

"Peter McHugh"

Peter McHugh recently graduated summa cum laude with a 3.9 average and a major in physics from one of New England's top-ranked private colleges. His scores on the Graduate Record Exam, as documented on score reports he sent us, were 650 on the GRE- Verbal, 750 on the GRE-Quantitative and 700 on the GRE Analytical. These scores correspond to the 91st, 89th and 86th percentile among all those who took the GRE between Oct. 1, 1994, and September 30, 1997 (ETS, 1998). In contrast, he scored only 82 on the MTT Reading test. Since the DOE has issued no technical documentation on the MTT, we cannot be certain what percentile ranking would correspond with a MTT Reading test score of 82; but on the basis of MTT scores we have collected, we estimate that Peter would fall somewhere in the range of 80th to 85th percentile on the MTT Reading test--obviously quite at odds with his performance on the GRE, especially since the population taking the GRE is more selective than that taking the MTT.

Even before graduation from college, Peter was offered a job to teach at one of the country's outstanding high schools in a state that recognizes a Massachusetts teaching certificate as certifying eligibility to teach.

Although a Rhode Island resident, Peter did practice teaching in Massachusetts schools and signed up to take the Massachusetts Teacher Tests in April 1998 in the belief that the results would not "count" toward certification. Less than two weeks before the test date, when he learned he would need to pass to receive a state teaching certificate, he threw all his energy into studying for the exam. Reviewing the test guide, he found one sample question for science, but nothing for physics, and, although he was told he would receive a list of test objectives, he never received one. In the absence of any guidance to what he might encounter on the exam, Peter took the state's curriculum frameworks and studied "about 15 hours a day" for two weeks, developing his own study. Like some other candidates interviewed, Peter found testing conditions for the listening portion of the test "absolutely atrocious" and the clarity of directions "not good at all." He explained:

> On the communication and literacy sections, sometimes you would have to write in the booklet and sometimes the answer sheet. You'd go back and forth. It wasn't consistent. It was a disaster. The proctors didn't understand. It ended up being a student who figured it out and explained it to the rest of us for one section. This took about five minutes and you started the test pretty frazzled.

Peter added, "I do remember leaving the test feeling that I didn't get to show I could read or write well." In the afternoon, Peter took the physics portion of the exam. This time, his concerns were less about testing conditions and more about test content. Specifically, he found content that is not mentioned in the Massachusetts curriculum frameworks and that, in any event, he considered inappropriate for high school physics. He reported:

> There was content on semiconductors. There were graphs and charts I was supposed to analyze, and I knew nothing about semiconductors. There were two 600-1000 word essay questions at the end. One was appropriate; the second was on motors and generators. Most high

> school textbooks have no more than a section on this. Until the morning of the test, I had never studied motors. In most high school physics courses, you don't get to this. The questions didn't accurately reflect what is covered in a high school physics course. The content went much beyond that. The last 25% of the test had content I had never seen before. I had to skip 7 or 8 questions because they were on concepts I had never seen in 14 college courses. I filled in the bubbles because there was no penalty for guessing.

Peter is currently teaching at the out-of-state high school that hired him in May. He teaches two "Honors" and three "AP" physics classes. He passed the literacy and communications portion of the MTT, but barely passed the physics portion of the MTT. He said, "If I'd been just a little lower, who knows where I'd be now. I wouldn't have this job."

"Allegra Karnofsky"

Allegra Karnofsky, a Connecticut resident, graduated in June 1998 from a nationally-known music college based at a private four-year university out of state. Before taking the Massachusetts Teacher Test in July, Allegra had passed the Reading, Writing and Math and Music Concepts and Content portions of the Praxis, with scores required for certification in Connecticut and New Jersey.

Allegra registered for the July test knowing little more than what she had read in newspapers about the April MTT results. She reported, "What was nerve-wracking was having heard that so many people failed in April." Although the NES registration booklet noted that study guides were available, Allegra's multiple calls to the testing company and Department of Education did not produce such a guide.

Because Allegra had taken the Praxis test for music teachers, she expected the MTT would test something about her teaching ability and was surprised at how little subject matter knowledge it covered. She said:

> There was a lot missing. There was some music theory, but I felt I was taking a test of music history. I am a K-12 certified teacher in Connecticut, and nowhere do you teach music history unless it's at college. The essay questions were very broad and open-ended, and both were music history -- for example, compare a 20th century composer and an 18th century composer. But that's not music education, it's music history. Most content was on classical, and there were one or two questions on American jazz. As a music education teacher, I need to know about a lot more than classical or jazz. There was no multicultural or world music, no Latin music.

Allegra added, "It wasn't geared toward a teacher. That was what troubled me the most. The past two years have been devoted to music education, not music history."

Prior to the test, Allegra had heard about the dictation exercise on the MTT Writing test; nonetheless, she found it "terrible." She said, "The tape recording was the worst experience of my life. The tape recorder the proctor was given was poor quality. It was muffled." She added, "The dictation is strange in its own way. As a teacher, you don't have to take down what your students say."

Allegra's passing scores on the Praxis allowed her to become certified to teach in New Jersey, where she had done practice teaching, and in Connecticut, and she believes new teachers should have to pass a standardized test in order to teach. She said:

> I think all teachers should know their subject. In the Praxis test I took, you take one test, then schedule the computer-based testing when you wanted. I don't see what's wrong with the Praxis test. If Massachusetts went for the computer-based test, it would make so much more sense instead of taking a paper-pencil test.

Allegra passed the MTT Reading test with a score of 74, failed the MTT Writing test by one point, and with a score of 50 failed the Music portion of the MTT by 20 points. Although she received tickets for a retest, the tickets came with no cover letter or explanation, and she had to call the Department of Education to ask whether she had to pay again for the retest.

At the time of the interview, Allegra was teaching music at a private school in Massachusetts. Saturday classes have precluded her re- taking the Massachusetts Teacher Test. When she called the Department of Education to ask whether the tests would be given at a time other than Saturday, she was told that alternative provisions could be made for religious reasons only. In February 1999, Allegra will begin work as a long- term substitute for kindergarten through sixth grade at a public school in Connecticut and look for a full-time job in that state. She says, "I don't know if I'm interested right now in retaking the [Massachusetts] test. I can just go on my way in Connecticut."

Summary of findings from interviews

The two vignettes above recount how two individuals experienced the MTT. These are only two of the 15 teacher candidates we interviewed. We realize fully that a sample of 15 self-selected individuals provides an extremely limited base from which to try to generalize to the experiences of all candidates who have taken the MTT. But we think it useful nevertheless to summarize some of the themes that emerged from interviews as possible causes for the low reliability and poor validity of the MTT. (A more detailed account of what was learned from the interviews appears in appendix 3.)

Many of our interviewees were dissatisfied with the information available about the MTT. Among items mentioned were the lack of a study guide, confusion over whether the April results would "count" towards certification, lack of information about conditions of retesting, and lack of detailed feedback on strengths and weaknesses of initial test performance.

A second theme among our small sample of interviewees was the conditions under which the MTT were administered. Most candidates interviewed found the general testing environment to be reasonable or on par with that of other tests taken. Nonetheless, close to half of them expressed concerns about the clarity of directions during MTT administration and about the conditions under which they performed the dictation exercise portion of the test.

Virtually all candidates interviewed mentioned the length of the MTT tests overall as excessive and believed that their 8-hour duration adversely affected their performance. Many compared the length of the MTT unfavorably with other tests they had taken and noted that the amount of writing required led to fatigue.

Teacher candidates interviewed also raised questions about the match between their "real-world" literacy skills and the test content in both the literacy and subject matter portions of the MTT. Regarding the former, candidates questioned the value of the dictation exercise and of specific questions, such as "Define a verb." As for the subject matter tests, interviewees expressed doubts about whether the tests matched Massachusetts curriculum frameworks and whether the test content was a reasonable reflection of the demands of real-world teaching. Some candidates interviewed also reported surprise that the MTT did not cover content they expected based on their experience taking other teacher tests, such as the Praxis or other states' teacher certification tests. Despite the range of concerns interviewees expressed, all agreed it was reasonable to ask teacher candidates to pass a test prior to certification. They were aware, as one put it, that "most of the professions have a test" and viewed testing one of the rites of passage into a profession. But in general, interviewees reported that the MTT compared unfavorably with other teacher certification tests they had taken.

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

Conclusions and Recommendations

The Ad Hoc Committee was formed in the summer of 1998 out of concern that important decisions were being based on the Massachusetts Teacher Tests (MTT) scores before any reasonable evidence had been produced concerning their reliability and validity. Since the DOE and NES have not made available any documentation on the reliability and validity of the MTT, in clear violation of professional standards concerning testing, and despite repeated requests for such documentation, the Ad Hoc Committee set out to study the technical merits of the new tests. Our original idea was compare individuals' scores on the MTT with scores on post-collegiate tests (such as the Praxis and the GRE) on which technical documentation is available. Toward this end we invited people to send us score reports on both the MTT and other tests.

As of December we had not received sufficient data to undertake a concurrent validity study, comparing MTT scores with those on established tests. But, in the meantime, we examined the reliability of the new tests. Specifically, using data on over 200 individuals who took the MTT in April and July 1998 (generously provided to us by eight institutions of higher education in the Commonwealth), we studied the test-retest reliability of the MTT. We found the correlations between April and July test to be extraordinarily low: about 0.30 for both the MTT Reading and Writing tests. Test-retest correlation coefficients for well-developed standardized tests typically range between 0.80 and 0.90. To examine the possibility that very low correlations were due to restriction of range (only people who scored below 70 on the April tests had to retake them), we corrected for attenuation due to restriction of range and estimated test-retest correlations for the unrestricted population of test-takers. The results indicated test-retest correlations of 0.50 to 0.70 -- still well below the reliability of well-developed tests.

We used these results to estimate the error of measurement in MTT scores. We found that MTT scores contain unusually high levels of measurement error--with an error of measurement on the new tests in the range of 9 to 17 points. We estimate that MTT Reading and Writing test scores contain two to three times the degree of error as well-developed tests.

Next, we compared pass and failure rates on the April and July administrations to consider the rates of misclassification on the MTT. Using both our test-retest sample, and a much larger sample of data reported on the DOE web site, we found that the MTT tests have very high rates of misclassification--as indicated by the fact that among those who "failed" either the MTT reading or writing test in April, more than 50% "passed" the test in July. Evidence suggests also that a fair number of people who "passed" the MTT did so simply because of error in the tests.

We also considered the content and construct validity of the MTT tests. At least one portion of the MTT Writing test (the dictation exercise) raises doubts about the content validity of the MTT and specifically their job-relatedness. Moreover, when we examined the correlation between MTT Reading and Writing test scores, the resulting correlations of about 0.50 raise serious doubt about their construct validity. Previous research suggests that the scores for tests of two related verbal constructs correlate in the range of 0.65 to 0.80.

Finally, we report on results of interviews with 15 candidates who took the MTT in April, July, or October (7 of whom passed and 8 failed). Since this was a small and self-selected sample, results are merely suggestive. But they indicate that the unreliability and poor validity of MTT scores may result from the lack of a study guide for the new tests, confusion over whether the April results would "count" towards certification, poor conditions of administration (in at least some test sites), simple fatigue resulting from the 8-hour duration of the tests, and test content. Although all those interviewed supported the idea of certification testing for teachers, as is common with other professions, many compared the MTT unfavorably with other teacher certification tests they had taken (e.g. the Praxis or NTE and certification tests in other states).

Recommendations

If the Commonwealth wants high standards for its teaching force, it must use assessments that meet similarly high professional standards. The current Massachusetts Teacher Tests fail to meet this criterion. Results from the April and July administrations of the MTT reveal that the new tests are so unreliable and of such poor validity that they are passing candidates who lack the knowledge and skills the MTT are allegedly testing and failing many who do have these skills. Therefore, the Ad Hoc Committee recommends that:

1. **The Massachusetts Board of Education immediately suspend administration of the MTT.** No exam at all is better than an unreliable exam that may be mistakenly failing 50% of qualified pre- service teachers while passing unknown numbers of unqualified ones.
2. **The Commonwealth convene an independent panel of testing experts to audit the development, administration and use of the MTT in light of both of professional standards for testing and the requirements of the Education Reform Act**. These experts should issue a report evaluating how well the first four administrations of the MTT meet accepted professional standards. If they find that the MTT fails to meet these standards, they should propose other approaches that will contribute to high-quality teaching in the Commonwealth.
3. **An investigation be launched into how and why the state has**

> **allowed the new MTT tests to be used.** An independent investigation into this matter is essential, since even before contracting with NES to develop the MTT, the DOE knew that a federal court had found that same firm to have "violated the minimum requirements for professional test development" with its teacher certification tests for Alabama. That the DOE nevertheless proceeded to allow the new MTT tests to be used, in obvious violation of professional standards on testing, to make important decisions about individuals before the validity and reliability of the new tests had been documented, was a course of action so imprudent as to call out for independent scrutiny.

As James Madison wrote in 1787, in the passage candidates were asked to transcribe in the April 1998 version of the MTT, "No man is allowed to be a judge in his own cause because his interest would certainly bias his judgment and, not improbably, corrupt his integrity." So too with organizations; the DOE, having implemented new teacher certification tests of undocumented validity and reliability, should not be allowed to judge its own cause.

Notes

1. In September 1998, the Massachusetts Department of Education (DOE) announced that the name Massachusetts Teacher Tests (MTT) was being changed to Massachusetts Educator Certification Tests (MCET), to reflect the fact that not just teachers but also other professional educators, such as counselors and principals, would be required to pass the new exams. However, throughout this report we refer to the Massachusetts Teacher Tests (MTT), since that is how they are most widely known.
2. These test standards have been developed by the American Educational Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurement in Education (NCME).
3. Richardson v. Lamar County Bd. of Educ. 729 F. Supp 806, M. D. Ala. 1989, p. 821. A portion of this decision appears in appendix 2 of this report.
4. NES President William Gorth signed the contract on February 23, and then Commissioner of Education Robert Antonucci on February 26, 1998.
5. The most common approaches for estimating internal consistency are the Cronbach alpha and split-half techniques.
6. We are submitting this report for publication and will make available to other investigators the complete set of data on which our reliability analyses have been based, but with the identities of the institutions of higher education removed.
7. In the remainder of this section of this report, we focus on MTT reading and writing scores. Among the more than 200 candidates whose MTT scores we obtained, there were many different

subject matter tests represented. Hence the sizes of samples for any one subject matter test were much small than those for the reading and writing tests.

8. Here we should explain why we devoted considerable attention to these anomalous "outlier" scores. Such unusual cases can have a disproportionate impact on summary statistics, such as means, standard deviations and correlation coefficients. Deletion of one or two extreme cases can change the summary statistics. Hence, as we explain below, we report reliability estimates not only for our entire test-retest sample, but also for a trimmed sample from which outlier cases have been deleted. Therefore, in summary
9. Candidates whose experiences are described in these vignettes have given their consent to these descriptions. We note, however, that specific details of their cases have been altered to protect their confidentiality.

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

References

American Educational Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurement in Education (NCME) (1985). *Standards for Educational and Psychological Testing.* Washington, DC: APA.

Anastasi, Anne (1976). *Psychological testing* (4th Edition). New York: Macmillan.

Conrad, Linda, Trismen, Donald and Miller, Ruth (Eds.) (1977). *Graduate Record Examination technical manual.* Princeton, NJ: Educational Testing Service.

Cronbach, Lee (1977). *Essentials of psychological testing* (3rd edition). NY: Harper & Row.

Donlon, Thomas (1984). *The College Board Technical Handbook for the Scholastic Aptitude Test.* New York: College Entrance Examination Board.

Educational Testing Service (1998). *GRE 1998-1999 Guide to the Use of Scores.* Princeton, NJ: Educational Testing Service.

Hambleton, Ronald K. (1999). Politicians fail, not the teachers. Education Connection (Winter issue), pp. 19-22.

Haney, W., Ludlow, L., Raczek, A., Stryker,S. and Jones, A. (1994). *Calibrating Scores on Two Tests of Adult Literacy: An Equating Study of the Test of Adult Literacy Skills (TALS) Document Test and the Comprehensive Adult Student Assessment System (CASAS) GAIN Appraisal Reading Test (Form 2)* . (Report prepared for the Manpower Demonstration Research Corporation) Chestnut Hill, MA: Boston College.

Haney, W., Madaus, G. and Kreitzer, A. (1987), Charms talismanic: Testing teachers for the improvement of American education. In E. Rothkopf (Ed.) *Review of Research in Education.* Volume 14, pp. 169-238.

Haney, W. & Raczek, A. (1993) *Surmounting outcomes accountability in education.* Washington, DC: U.S. Congress Office of Technology Assessment.

Hart, Jordana (1998). "The teachers test: Madison, via Silber." Boston Globe, 8/5/98, pp. A1 & A20

Jackson, Bailey (1998). Letter dated September 18, 1998 to Dr. David P. Driscoll, Commissioner of Education (on behalf of the Commonwealth Education Deans' Council).

Linn, Robert (1983). Pearson selection formulas: Implications for studies of predictive bias and estimates of educational effects in selected samples. *Journal of Educational Measurement*. 20:1, 1-16.

Lord, F. and Novick. M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Madaus, G. (1998). Testimony on the testing portion of Governor's Bill (H. 5677). (Testimony before the Education Committee, Massachusetts House of Representatives. July 15, 1998).

Pressley, Darrell S. (1998). "Dumb struck: Finneran slams 'idiots' who failed teacher tests." Boston Herald, 6/26/98 pp. 1, 28.

Thorndike, Robert and Hagen, Elizabeth (1977). *Measurement and evaluation in psychology and education* (4th Edition). New York: Wiley.

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

Appendix 1
The Massachusetts Teacher Tests: A Chronology

- **7/85** The Massachusetts legislature passed a bill that, among other things, required candidates for teacher certification to pass a "standardized exam in his or her subject field [and] a standardized exam of communication and language skill" (MGL, 1985, Chapter 188, Section 15).

- **6/93** The Massachusetts legislature passed the Education Reform Act of 1993 which included language that required candidates to pass tests, specifically "a writing and subject matter test," in order to be certified.

- **10/96** The Board of Education had "an initial discussion about implementing ... a two-part test for teacher certification .... The Commissioner [Robert Antonucci] recommended that the Board set October 1, 1997 as the implementation date. The Board agreed to discuss this further and take action at the November meeting" (Massachusetts Department of Education, Board in Brief, 10/24/96).

- **11/96** The Massachusetts Board of Education "voted to endorse a recommendation by Commissioner Antonucci to require all candidates for teacher certification as of January 1, 1998 to pass a standardized test in communications and literacy skills and subject matter knowledge" (Massachusetts Department of Education, Board in Brief, 11/21/96). The Board also voted that Commissioner Robert Antonucci should "proceed at once with the selection of a test vendor with the aim of having the test available for review by the Board no later than October 1, 1997" (Massachusetts Department of Education, Videotape of Board Meeting of 11/18/96).

- **2/97** The DOE issued on 2/24/97 a Request for Responses (RFR) from prospective teacher certification test contractors. The RFR stated that vendors should describe in their bid 1) how they would deliver a "technical report to the Department of Education following the use of each new form of the tests with a summary for public dissemination" (p. 15); and 2) their "plan for consultation with a technical committee of nationally recognized experts recommended by the Contractor (external to the Contractor's organization) ... The technical committee will review the test items, test administration, and scoring procedures for validity and reliability and report its findings to the Department of Education" (p. 11). The RFR also asked bidding vendors to meet a timetable that included the following critical dates and events:

- **5/97-11/97** Advisory committees review sample test questions, scoring and evaluation criteria, and plan for quality control. Pilot tests conducted.

- **11/97** Test materials submitted to DOE for approval.

- **1/98-6/98** Monitoring of test quality and standardization; reliability study due.

- **12/97** Commissioner Antonucci reported to the Board on 12/18/97 that he had selected National Evaluation Systems (NES) to develop and administer the new tests, but some issues needed to be settled before executing a contract: "Issues still under discussion with the contractor include the test administration schedule and, more importantly, ensuring that the test will be rigorous and of high quality, based on college-level content" (Massachusetts Department of Education, Board in Brief, 12/15/97).

- **1/98** The DOE released a registration bulletin and an informational packet concerning these tests. The informational packet, titled "Massachusetts Teacher Tests, Questions and Answers, January, 1998" stated the following:

  > "**[Question]**: If I am now enrolled in a teacher preparation program, when should I take the tests?
  > **[Answer]**: Candidates who expect to complete their teacher preparation programs by August 31, 1998 are encouraged to take the teacher tests on either April 4 or July 11, 1998. Candidates who take the tests on these dates will satisfy the testing requirement automatically. Candidates who take the tests beginning with the October 3, 1998 administration will be required to achieve a qualifying score in order to be certified" (p. 3).

  The official 1998-1999 Registration Bulletin also informed candidates that "no qualifying score will be established until after the first two administrations of the tests ... Candidates who must take the tests and are eligible to participate in those first two administrations will satisfy the testing requirement by completing the tests" (p. 2).

- **1/98** The DOE mailed content validation surveys to school districts and teacher preparation programs asking teachers and professors to review and comment on test objectives for 31 different tests. Respondents were asked to complete and return these surveys by January 31, 1998. 4,300 eligible respondents replied.

- **2/98** Educators participated in test validation conferences for the MTT held on 2/10 and 2/12/98. Participants reviewed test bank items for: 1) match of item to test objective, 2) accuracy, 3) freedom from bias, and 4) job-relatedness (for Massachusetts teachers).

- **2/98** Robert Antonucci resigned the position of Commissioner and was later replaced by Frank Haydu as Interim Commissioner.

- **2/98** Approximately 1,500 college juniors and seniors in the state's teacher preparation programs participated in pilot testing of new open-ended and multiple-choice questions for selected tests during the early part of February.

- **2/98** The DOE and NES signed a contract to have NES develop the MTT. William Gorth signed for NES on 2/23/98; exiting Commissioner Robert Antonucci signed for the DOE on 2/26/98.

- **3/98** The DOE issued a "Study Guide" that stated the following: "This is a preliminary edition of the Massachusetts Teacher Tests study guide. An expanded set of study guides, including sample questions and open-ended questions from each test field, will be available beginning in August 1998."

- **3/98** The DOE withdrew the study guide because the Written Mechanics Exercise, designed to assess candidates' knowledge of spelling, punctuation and capitalization, was changed. Originally, candidates were to be asked to fill in (three to five) missing portions of six different sentences that were printed in the test booklet. They were to do this as an audiotaped narrator read each sentence, several times over, in its entirety. For the April exam, however, candidates were asked to transcribe a 156-word text written by James Madison in 1787. This text, part of the Federalist Papers, was read three times by an audiotaped narrator. **(See copy of this text at the end of this chronology.)**

- **3/98** In a reversal of previous policy, the DOE announced on 3/25/98 that eligible candidates taking the April and July tests would no longer qualify automatically; instead, they would have to achieve a passing score to be provisionally certified.

- **4/98** The first round of tests were administered on 4/4/98.

- **4/98-7/98** Scoring panels met to recommend cut scores for every test.

- **6/98** The Board of Education voted on 6/22/98 to set the cut score at 1 standard error of measurement below the scores recommended by the scoring panels.

- **7/98** The Board met again on 7/1/98, at the request of Acting Governor Cellucci, and voted to raise the cut score to the level originally recommended by the scoring panels. During this meeting, Frank Haydu resigned the position of Commissioner of Education. David Driscoll was later named Acting Commissioner.

- **7/98** NES mailed score reports on 7/6/98 to April test-takers.

- **7/98** The second round of tests were administered on 7/11/98. The Ad Hoc Committee distributed flyers at five of the six test sites.

- **9/98** The DOE released a "Test Information Booklet" that contained a) sample questions from each section of the communications and literacy test; b) one sample multiple-choice question for just thirteen of the (43) subject exams; c) one sample open-response item for just one subject exam; and d) one sample oral expression and one sample written expression item from one (of the 7) foreign language exams.

- **9/98** The DOE announced that as of 9/30/98 the title of the educator certification testing program changed from Massachusetts Teacher Tests to Massachusetts Educator Certification Tests.

- **10/98** The third round of tests were administered on 10/3/98. The Ad Hoc Committee distributed flyers at all six test sites.

- **1/99** The fourth round of tests were administered on 1/9/99. The Ad Hoc Committee distributed flyers at all six test sites.

Candidates who sat for the first administration of the MTT on April 4, 1998, were asked to transcribe the following text written by James Madison in 1787. This exercise, which constituted the Written Mechanics section of the communications and literacy test, is intended to assess test-takers' knowledge of capitalization, punctuation, and spelling. An audiotaped narrator read the full text three times:

> "No man is allowed to be a judge in his own cause because his interest would certainly bias his judgment and, not improbably, corrupt his integrity. With equal, no, with greater reason, a body of men are unfit to be both judges and parties at the same time. Yet, what are many of the most important acts of legislation but so many judicial determinations, not indeed concerning the rights of single persons, but concerning the rights of large bodies of citizens? And what are the different classes of legislators but advocates and parties to the causes which they determine? It

is in vain to say that enlightened statesmen will be able to adjust these clashing interests and render them all subservient to the public good. The inference to which we are brought is that the causes of faction cannot be removed, and the relief is only to be sought in the means of controlling its effects."

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

~ EPAA Vol. 7 No. 4: Haney, Fowler, Wheelock, Bebell & Malec "Massachusetts Teacher Test" ~

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

Appendix 2
Richardson v. Lamar County Bd. of Educ. 729 F. Supp 806 (M. D. Ala. 1989) (Excerpted)

United States District Court, M.D. Alabama, Northern Division.

Nov. 30, 1989.

MEMORANDUM OPINION

MYRON H. THOMPSON, District Judge.

Plaintiff Alice Richardson, an African-American, has brought this lawsuit claiming that defendant Lamar County Board of Education [FN1] wrongfully refused to renew her teaching contract in violation of Title VII of the Civil Rights Act of 1964, as amended. [FN2] Richardson charges the school board with two types of discrimination under Title VII. First, she asserts a claim of "disparate treatment": [FN3] that the school board refused to renew her contract because of her race. Second, she asserts a claim of "disparate impact": that the board's stated reason for not renewing her contract--that she had failed to pass the Alabama Initial Teacher Certification Test--is impermissible because the test has had a disparate impact on African-American teachers. The court's jurisdiction has been properly invoked pursuant to 42 U.S.C.A. § 2000e-5(f)(3).

FN1. Richardson has sued not only the Lamar County Board of Education but also its superintendent and members. However, because Richardson may obtain full relief from the school board the court has not treated the board members and the superintendent separately from the school board.

FN2. Title VII is codified at 42 U.S.C.A. §§ 2000e through 2000e-17.

FN3. Richardson's disparate treatment claim is also based on 42 U.S.C.A. § 1981 and the fourteenth amendment, as enforced by 42 U.S.C.A. § 1983, Jett v. Dallas Independent School District, 491 U.S. 701, 109 S.Ct. 2702, 105 L.Ed.2d 598 (1989), with jurisdiction premised on 28 U.S.C.A. §§ 1331, 1343. Because a plaintiff must prove intentional discrimination to establish a disparate treatment claim under § 1981, § 1983 and the fourteenth amendment as well as under Title VII, Stallworth v. Shuler, 777 F.2d 1431, 1433 (11th Cir.1985), and because Richardson is seeking the same relief under all these statutory provisions, the court need not address separately her theories under §§ 1981, 1983, and the fourteenth amendment. The court also need not address whether Richardson has stated a cognizable claim under § 1981. Patterson v. McLean Credit Union, 491 U.S. 164, 109 S.Ct. 2363, 105 L.Ed.2d 132 (1989).

Based on the evidence presented at a nonjury trial, the court concludes that Richardson may recover on her disparate impact claim but not on her disparate treatment claim. The court's disposition of Richardson's disparate treatment claim is simple and direct. The court simply applies the procedure set forth by the Supreme Court in Texas Department of Community Affairs v. Burdine, 450 U.S. 248, 101 S.Ct. 1089, 67 L.Ed.2d 207 (1981). The court's disposition of her disparate impact claim is, however, much more difficult. The court first addresses and finds meritless two defenses raised by the school board: that Richardson's disparate impact claim is barred by principles of collateral estoppel and res judicata; and that under the framework set forth in Price Waterhouse v. Hopkins, 490 U.S. 228, 109 S.Ct. 1775, 104 L.Ed.2d 268 (1989), Richardson would not have been reemployed even if she had passed the state certification test. The court then goes through a lengthy application of the disparate impact analysis outlined by the Supreme Court in Wards Cove Packing Co., Inc. v. Atonio, 490 U.S. 642, 109 S.Ct. 2115, 104 L.Ed.2d 733 (1989).

I. BACKGROUND

Richardson taught in the Lamar County School System for three years, from 1983 to 1986. She was, however, unable to obtain a permanent teaching certificate and therefore had to teach with temporary and provisional certificates. To obtain a permanent certificate, Richardson, like all other teachers in the state at that time, had to *809 pass the Alabama Initial Teacher Certification Test, which consisted of a "core" examination and an examination aimed at the specific area in which the teacher sought to teach. Richardson wanted to teach in the areas of early childhood education and elementary education, and thus could meet the certification test's specific area requirement by passing the examination in either area. Between 1984 and 1986, Richardson failed the early childhood education examination twice and the elementary education examination three times.

In the spring of 1986, the Lamar County Board of Education decided that the elementary school where Richardson taught should be consolidated with another school. Because fewer teachers would be needed, the school board informed 15 nontenured teachers, including Richardson, that their contracts would not be renewed for the 1986-87 school year. Four of the 15 teachers were, however, rehired. Richardson, who would have acquired tenure if she had been rehired, was not one of the four.

Approximately a year later, in May 1987, this court enforced a consent decree requiring the State Board of Education to issue permanent teaching certificates to a court-defined class of black teachers who had failed the state teacher certification test. [FN4] Richardson received her certification pursuant to the consent decree.

FN4. Allen v. Alabama State Board of Education, 816 F.2d 575 (11th Cir.1987) (directing district court to enforce consent decree); Allen v. Alabama State Board of Education, Civil Action No. 81-697-N (M.D.Ala. May 14, 1987) (enforcing the consent decree).

[NOTE: Omitted from this reproduction of Judge Thompson's opinion are several pages in which he discussed: II. DISPARATE TREATMENT CLAIM; and III. DISPARATE IMPACT CLAIM. The remainder of the opinion is reproduced in its entirety.]

[15] Since Richardson has established that the early childhood education and elementary education examinations had an adverse racial impact, the burden shifts to the Lamar County Board of Education to produce evidence of employment justification. An understanding of the history of the Alabama Initial Teacher Certification Test is important to determining whether the school board has met its burden and, if so, whether Richardson has, in turn, shown that the school board's justification for the certification test has no basis in fact.

a. History of the Early Childhood Education and Elementary Education

Examinations

In 1979, amidst a national groundswell in favor of teacher competency testing, the Alabama State Board of Education placed development of a uniform certification test at the head of its agenda. It retained a professor at Auburn University to conduct a feasibility study regarding implementation of a teacher testing program in Alabama; the state's Assistant Superintendent for Teacher Certification also participated in the study. After a rather cursory investigation, the two educators recommended implementation of a testing program similar to one designed by a private test developer for the State of Georgia.

The State Board agreed with the recommendation. In January 1980, it awarded a contract to the private test developer on a noncompetitive basis. [FN29] While the board did not always express its purpose for imposing the test requirement with perfect clarity, both the test developer and the board understood that the test would measure whether a teacher possessed enough minimum content knowledge to be competent to teach in the classrooms of Alabama.

FN29. Board members anticipated that the test requirement would adversely impact against African-American applicants for teaching certificates. However, the same decision would have been reached without consideration of that factor. The board's action was predicated on a legitimate concern for improving the quality of education in Alabama.

The time frame for development of the Alabama Initial Teacher Certification Test, as it came to be known, was quite short. The test developer had one year to complete development and implementation of 36 separate examinations. The test developer created a "core" examination and 35 additional examinations that covered specific subject areas. As stated, a teacher had to pass the core examination and one subject area examination in order to receive certification.

The Assistant State Superintendent, the sole ranking state official charged with oversight of the private test developer's contract compliance, had a doctorate in educational administration; but neither he nor anyone on his staff had any expertise in test development. And no outside experts were retained to monitor the test developer's work. The developer's work product was accepted by the state largely on the basis of faith.

The test developer began by preparing a preliminary planning document. It next asked the State Department of Education to appoint Alabama educators to the various committees and panels necessary for completion of the project. According to criteria provided by the developer, these educators were selected to represent a fair cross section of persons from different geographic areas throughout the state. They were also selected in such a way that African- Americans and women were fairly represented overall; however, not all committees and panels had minority representatives.

The test developer's technical staff and subject area consultants then formulated topic outlines for the various examinations. They consulted state education standards, state courses of study, materials related to Alabama's student competency tests, and examples of textbooks used in Alabama public schools. They also developed actual test objectives. These objectives were more explicit statements of concepts embedded in the topic outlines. The objectives were reviewed by the developer's editors and management. The developer's in-house work was far below average.

*818 In October of 1980, approximately 200 Alabama educators attended a two-day conference to review the topic outlines and objectives for 36 examinations. They had previously been mailed orientation materials. After additional orientation, they were divided into curriculum committees to review the topic outlines for comprehensiveness, organization, accuracy, and absence of bias. The committees then reviewed the objectives to ensure that they matched the topic outlines. Taxonomic level, significance of content, accuracy, level of specificity, suitability, and lack of bias were considered. Decisions were reached by consensus during both stages of review. Modifications and deletions were recorded by the test developer's personnel assigned to each committee. In some cases, however, the developer made additional changes, or ignored suggested changes, without obtaining clearance from committee members. No effort was made at any time to link the topic outlines and objectives to the state-mandated curriculum for teacher training programs.

The test developer then sent a job analysis survey packet to approximately 3,000 in-service teachers throughout Alabama. The purpose of this survey was to determine the job relatedness of the test objectives. [FN30] However, in nine fields where there were fewer than 200 teachers throughout the state, the test developer's process resulted in very small response rates. The survey packet was sent to persons certified and teaching in specific content areas. The packet included a set of objectives for that content area, a survey form, and a set of instructions. The teachers were asked whether they had taught or used each objective in the past two school years. If the answer was yes, they were asked to rate the objective in terms of time and essentiality. The scales used to record those responses were balanced in favor of indicating that an objective was job related, and teachers were instructed to resolve doubts in favor of job relatedness. The results of the job analysis survey were tallied in such a way that responses from only those who indicated that an objective had been used in the last two years were reflected in the data. Those who indicated that an objective had not been used were ignored.

FN30. A stratified random sampling technique was employed to select survey respondents and a fair cross section of teachers was generally achieved.

In January of 1981, the curriculum committees met for a second time. They were provided results from the job analysis survey and were asked to determine which objectives should generate questions to appear on the examinations. This step was called "objective selection." The survey results were a major determinant of which objectives were ultimately selected.

The test developer then prepared a "blueprint" for each examination. These blueprints specified the number of test questions, or items, necessary to measure each objective. Test items were drafted by the test developer's content area consultants and edited by its staff. Again, the developer's in house work was far below average.

In March of 1981, the test items were reviewed by Alabama curriculum committees for "item/objective" match, significance of content, accuracy, clarity, and absence of bias. This "item review" process lasted for two days. Committee revisions were recorded by the test developer's personnel. However, in some cases, the developer ignored the suggested changes, or made additional recisions, without consulting committee members for approval. In other cases, the developer simply added new items that had never been reviewed by committee members. As many as 20 items for each 120-question examination fell into one of these categories.

In late April of 1981, the test developer convened a separate group of educators to review the test items once again for content validity. The purpose of this session, which lasted one day, was to provide an independent check against the judgments already rendered by the previous committees of Alabama educators. The new panelists reflected a fair cross section of persons in their field and were qualified to make content validity judgments in their *819 field. Each educator worked separately, but votes were tallied as if educators had served on a committee. After orientation, the educators were asked to judge whether each item matched its objective, was accurate, was free of bias, and was not tricky, misleading, or ambiguous. If the item met these criteria, the item was rated content valid by that judge. If the item was deemed invalid, the judge's reason for rejecting that item was recorded. The test developer compiled these content validity ratings; a level of agreement among judges greater than 50% was required for an item to be deemed content valid. While a majority of items appearing on the final test instruments reflected the judgment of Alabama educators that those items were content valid, a significant number of items appearing on the tests did not reflect that judgment. These included those items that had been revised by the developer without obtaining clearance from the panelists. [FN31]

FN31. The test developer did not convene separate panels of minority educators at any stage of the item review or content validity process to screen items for possible bias.

The judges were also asked to make cut-score decisions for those items they had rated content valid. For these items, and those items only, judges were asked whether a teacher with minimum content knowledge in the field should be able to answer the item correctly. A yes-no response was requested. Judges were disqualified from making that same cut-score determination for any item they had previously rated content invalid. In essence, their expert judgment as to those items was ignored.

The test developer then assembled and produced the actual test instruments for all 36 examinations. Each examination had 100 items tentatively designated as scoreable and 20 items tentatively designated as nonscoreable. The examinations were first administered to a group of actual candidates. The test developer had originally contemplated a separate field tryout, but time constraints prohibited such a course. After the first administration, the developer examined item statistics to flag problem questions. Based on this item analysis, it selected 100 scoreable items and 20 nonscoreable items for each examination. The developer did not conduct empirical bias studies to determine whether the difficulty of items varied according to the race of examinees.

The test developer then set a minimum cut score for each examination. The developer's original plan was to take the panelists' cut-score ratings and subject them to a 10% non-cumulative binomial algorithm.

This level of agreement among judges would then determine the minimum cut score. However, the developer's procedure yielded cut scores that were so astoundingly high that they signaled, on their face, an absence of correlation to minimum competence. For example, of the more than 500 teachers who took the first administration of the core examination, none would have passed if the original cut-score methodology had been followed.

Faced with this problem, the test developer made various mathematical "adjustments" to the original cut score. First, the developer applied a 10% cumulative binomial algorithm. When the cut scores still remained too high, it applied a 5% cumulative binomial algorithm. This process of applying successively stricter algorithms was referred to at trial as a "binomial twist." The developer engaged in this process without consulting the State Department of Education or any Alabama educators. In two fields--that of Music and that of Speech, Communication, and Theatre--the 5% binomial twist yielded cut scores that were much too low. The developer simply applied a different mathematical algorithm to those examinations; again, the developer consulted no one. For all special education and school counseling examinations, the developer recommended a uniform cut score cap of 80 to the State Department of Education. This recommendation was based on the developer's experience in the Georgia testing program. However, in Georgia, the decision to place a cap on cut scores was reached by state officials in conjunction with Georgia educators. [*820 FN32]

FN32. Although the cut scores in the special education area were intended to serve as an upper limit, the cut scores on five of those examinations were actually raised to 80 to achieve uniformity.

The State Department of Education was then given the option of dropping the cut scores, as set by the developer, by two or three standard errors of measurement (SEM's). It was clear at that time that cut scores, even after the various adjustments catalogued above, were not measuring competence. For example, even after the developer's 5% binomial twist, 78% of the teachers taking the first administration of the core examinations would have failed. The same would have been true for 93% of those taking the school counseling examination, 89% of those taking the learning disability examination, and 97% of those taking the library media examination. Instead of challenging what the developer had done, the state simply dropped the cut scores three SEM's in order to arrive at a "politically" acceptable pass rate. In so doing, the state knew that the examinations were not measuring competency.

In 1982, the test developer formulated nine additional examinations. Its test construction procedures and quality of execution were essentially the same, with the following exceptions. First, the developer's job analysis survey form contained a rating scale with additional errors. Second, a more restrictive binomial table was used to calculate agreement among panelists on content validity questions. Third, a more accurate cut-score methodology was employed.

In 1983, the developer conducted a "topicality review" to update ten of the examinations already in use. A curriculum committee performed item and objective review. The committee's tasks were to determine whether items had become stale because of changes in the teaching field and to identify problems with items by reference to item statistics for the first eight administrations of the certification test. On average, 50% of the items in any given examination were replaced or revised. The developer did not convene a separate panel, as it had during the initial test development, to provide an independent screen for content validity, nor was an independent cut-score panel convened. The curriculum committee provided ratings used to set cut scores.

b. Validity of the Early Childhood Education and Elementary Education

Examinations

The Lamar County Board of Education contends that the state teacher certification test was designed to determine whether a teacher is competent to teach in Alabama's classrooms. Richardson claims, as stated, that the early childhood education and elementary education examinations were invalid, that they did not measure competency.

Generally, validity is defined as the degree to which a certain inference from a test is appropriate and meaningful. APA Standards at 94. [FN33] It is suggested that validity evidence must necessarily be restricted to success on the job; and, to be sure, there are Title VII decisions that have approached the question of validity by asking whether a given score on a test yields an appropriate and meaningful inference about successful performance on the job. See, e.g., Contreras v. City of Los Angeles, 656 F.2d 1267, 1271-1272 (9th Cir.1981), cert. denied, 455 U.S. 1021, 102 S.Ct. 1719, 72 L.Ed.2d 140 (1982); Guardians Association of New York City Police Dept., Inc. v. Civil Service Commission, 630 F.2d 79, 91 (2d Cir.1980), cert. denied, 452 U.S. 940, 101 S.Ct. 3083, 69 L.Ed.2d 954 (1981). However, there is no magic to using success on the job as an anchor point for validity. Success on the job is just one of many constructs that a test can measure. Thus, a sound inference as to a different construct, such as minimal competence, may also form the basis for a finding of validity. In short, a test will be valid so *821 long as it is built to yield its intended inference and the design and execution of the test are within the bounds of professional standards accepted by the testing industry. APA Standards at 9; cf. Washington v. Davis, 426 U.S. 229, 247 & n. 13, 96 S.Ct. 2040, 2051 & n. 13, 48 L.Ed.2d 597 (1976) (validity need not be limited to inference about success on the job).

FN33. The term APA Standards is a shorthand reference for the American Educational Research Association, American Psychological

Association, National Council on Measurement in Education, Standards for Educational and Psychological Testing (1985).

In order to be valid, a licensure or certification test must support the inference that persons passing the test possess knowledge necessary to protect the public from incompetents. APA Standards at 63. Part of an appropriate validation strategy for licensure and certification tests is to define clearly and correctly the domain of minimum content knowledge necessary for competence. The test domain, once defined, must then be translated into actual test questions that measure competence. At all stages, validity flows from the expert judgment of practitioners in the field being tested. The test developer's role is to employ professionally accepted practices that accurately marshal the expert judgment of those practitioners. When the questions on a given test actually measure what practitioners in the field consider to be content knowledge associated with competency, the test instrument is held to possess content validity. However, mere content validity does not alone establish test validity. No matter how valid the test instrument, an inference as to competence or incompetence will be meaningless if the cut score, or decision point, of the test does not also reflect what practitioners in the field deem to be a minimally competent level of performance on that test. Again, the test developer's role in setting a cut score is to apply professionally accepted techniques that accurately marshal the judgment of practitioners.

In assessing the overall validity of the Alabama Initial Teacher Certification Test, the court must therefore address both content and cut-score validity. The test developer retained by the State Board of Education followed a multi- step procedure to build 36 teacher certification examinations in 1981. With minor variations, it followed the same procedure when it built nine additional examinations in 1982. The developer then applied a third procedure when it revised ten examinations in 1983. The content validity of each of these examinations turns on whether the developer's procedures were adequate, or were outside the bounds of professional judgment. For reasons that follow, the court concludes that the developer's procedures violated the minimum requirements for professional test development. Accordingly, none of the examinations, including the early childhood education and elementary education examinations, possesses content validity.

The test development process was outside the realm of professionalism due to the cumulative effect of several serious errors committed by the developer when it formulated the 45 examinations in 1981 and 1982. First, while practicing teachers were asked to offer their judgment about the job relatedness of test objectives, it is clear that the test developer's survey instrument distorted that judgment. Scales were balanced in favor of finding job relatedness and respondents were specifically instructed to resolve all doubts in favor of job relatedness. Moreover, the response of those teachers who indicated that they had not used an objective was ignored.

Second, Alabama educators serving on curriculum committees selected test objectives based on those survey results. It has been suggested that the survey was used only in an advisory capacity and that any survey errors were offset by the overall judgmental process undertaken by committee members. However, it is plain that the survey was conducted to solicit critical firsthand knowledge from in-service teachers. It is equally plain that curriculum committee members, aware that the survey had been conducted for that purpose, took the survey results quite seriously. The court concludes that the overall judgmental process for determining job relatedness of test objectives was distorted significantly by survey error.

Third, a significant number of items appearing on the examinations failed to reflect accurately the collective judgment of curriculum committee members. In some *822 cases, changes to actual test items were not implemented. In other cases, items that had never been reviewed by a curriculum committee appeared on examinations. It is suggested that, in any testing program of this size, a certain number of errors of this type will be found. The court agrees with this proposition in principle; however, the evidence reflects that the error rate per examination was simply too high.

Fourth, Alabama educators were never asked to determine whether the test items themselves were job related, even though such an approach is standard practice in the testing industry.

Fifth, many items appeared on the examinations even after they had been rated content invalid by the requisite number of Alabama panelists. It is suggested that, before any such item appeared on a final test form, it was revised by the test developer, and that all revisions were approved by Alabama panelists. However, neither the State Board of Education nor the test developer produced any documentation of this alleged revision and approval process. Moreover, not a single panelist was called at trial to confirm that the process had actually occurred. The court finds that no such process occurred and that the test developer simply substituted its own judgment for that of Alabama educators.

In 1983, the test developer conducted a topicality review for ten of the examinations already in use. It is suggested that, even if those ten examinations were previously content invalid, they gained content validity by way of the topicality review process. The court does not agree. The topicality review process resulted in changes to, or replacement of, only about 50% of any given examination's 120 items. Items that were not revised or replaced therefore remained just as invalid as they were at birth. Moreover, as to items that were revised or replaced, there was no separate content validity determination. The court agrees with Richardson's experts that, on balance, these two factors rendered the ten examinations subjected to the 1983 topicality review to be content invalid as well. [FN34]

FN34. The court does not agree that the test developer's multi-step test development process was inherently self-correcting. There is substantial support in the record for the view that errors at one step not only survived the next step, but also created new errors.

Moreover, the fact that a validity study for the National Teachers Examination was upheld in United States v. South Carolina, 445 F.Supp. 1094 (D.S.C.1977), aff'd 434 U.S. 1026, 98 S.Ct. 756, 54 L.Ed.2d 775 (1978), does not mandate the same result here. The validity of the present examinations must be assessed on the basis of evidence now before the court. Cf. York v. Alabama State Board of Education, 581 F.Supp. 779, 786 (M.D.Ala.1986) ("tests are not valid or invalid per se ...; the fact that the validity of a particular test has been ruled upon in prior litigation is not necessarily determinative in a different factual setting").

Richardson advances an array of challenges to the cut-score methodology employed by the test developer. It is clear that, as to the 35 examinations developed in 1981, the cut scores bear no rational relationship to competence as that construct was defined by Alabama educators. [FN35] The *823 evidence reveals a cut-score methodology so riddled with errors, that it can only be characterized as capricious and arbitrary. There was no well- conceived, systematic process for establishing cut scores; nor can the test developer's decisions be characterized as the good faith exercise of professional judgment. The 1981 cut scores fall far outside the bounds of professional judgment.

FN35. The court must point out that three of Richardson's arguments with respect to the 1981 cut scores clearly lack merit. First, she asserts that Nassiff's 1978 "Two-Choice Angoff" method for yielding an original cut score was and is "without professional endorsement." However, professional literature published well before the initiation of Alabama's testing program endorsed methodologies similar to Nassiff's approach. See R. Thorndike, Educational Measurement at 514-515 (1971). Moreover, while current professional literature does not grant Nassiff's method the highest possible marks, it certainly does not condemn it as being wholly outside the bounds of professional judgment. See Berk, A Consumer's Guide to Setting Performance Standards on Criterion Referenced Tests, 56 Rev. of Educ. Research 137, 148 (1986). Second, Richardson contends that Nassiff's method was largely unproven and that an alternative cut-score methodology should have been used at the same time as a backup. While the court agrees that this might have been advisable, there is no evidence that the failure to use a backup cut-score method was unprofessional. Third, Richardson argues that the test developer's recent adoption of a more sophisticated cut-score methodology signals the bankruptcy of Nassiff's 1981 method. The court does not agree. The fact that, with new developments in the field, the test developer later changed its methodology should not be held against it as an admission of error.

First and foremost, it is undeniable that cut scores for the 35 examinations developed in 1981 do not reflect the judgment of Alabama educators who served as panelists on the minimum cut score committees. This is a crucial error, because competence to teach is a construct that can only be given meaning by the judgment of experts in the teaching profession. Here, expert panelists who rated an item invalid as to content were automatically disqualified from going on to indicate whether that item should be counted toward the minimum cut score. This means that when a panelist indicated that an item should be excluded--because it contained inaccurate content, did not measure an objective, was tricky, ambiguous, or misleading, or was biased--that panelist's opinion was ignored for purposes of determining whether the item measured competence and should contribute to the cut score. The exclusion of such opinions resulted in a series of cut scores that reflected a distorted notion of competence.

Second, the court has no doubt that, after the results from the first administration of those 35 examinations were tallied, the test developer knew that its cut-score procedures had failed. The proof of this fact is that none of the more than 500 teachers who took the first administration of the core examination would have passed if the original cut score, calculated according to the developer's original plan, had been utilized. The court cannot conclude that all Alabama teachers who took that examination were totally and completely incompetent. It follows, therefore, that the developer knew that its cut-score procedure had utterly failed to reflect a valid construct of competence.

Third, instead of notifying the State Department that its cut-score procedure had malfunctioned, the test developer attempted to mask the presence of system failure by making various unilateral mathematical "adjustments" to the original cut score until an "acceptable" score had been reached. The most common adjustment was application of a "binomial twist" to the data collected from Alabama educators. This adjustment tended to lower cut scores. It is argued that lowering cut scores offset any system failure that might have occurred previously. This argument, however, misses the mark. The critical factor with respect to cut-score validity is not whether there was a net change in cut- score level, but whether the cut score itself accurately reflected the expert judgment of Alabama educators about whether examinees possess the competence to teach. This construct of competence cannot be guessed at by out-of-state test makers. It is also argued that the developer's resort to the "binomial twist" was an exercise of "tempered judgment" in light of actual examination data. Again, however, the fatal error is that it was the developer, and not Alabama educators, that exercised this judgment. [FN36]

FN36. It is argued that the binomial twist was, in fact, implemented in consultation with the State Department of Education, and that such consultation somehow injects the judgment of Alabama educators into the cut-score process. However, the evidence is clear that the developer never consulted any official at the State Department of Education with respect to the binomial twist. In fact, the State Department was not advised of that twist until shortly before trial.

Fourth, in two fields--that of music, and and that of speech, communication and theatre--the 5% binomial twist yielded cut scores that were much too low. In those areas, the developer simply applied a different mathematical algorithm to arrive at an acceptable cut score. Again, the developer substituted its judgment about competence for that of Alabama educators.

Fifth, for all special education and school counseling examinations, a uniform cut score of 80 was adopted. To be sure, the *824 State Department of Education made this decision, based on a policy judgment that no score should exceed 80. However, it is clear that the developer played an advisory role in that decision and that its advice was completely irresponsible. The developer recommended holding the scores at 80 based on its experience in the Georgia testing program. However, in Georgia, the decision to place a cap on cut scores was reached by the State Department of Education in conjunction with Georgia educators. The test developer never suggested that the State Department consult Alabama educators, and there is no evidence that such consultations in fact occurred. In effect, the developer assumed that the judgment of Georgia educators in a different testing program would be good enough for the people of Alabama. Once again, cut scores bore no relation to the expert judgment of Alabama educators. Moreover, if the rationale for adopting a cut score of 80 was to place a cap on such scores, it is difficult to understand why the cut scores for five special education examinations were actually raised to 80.

Sixth, the State Board did not drop the cut scores, as set by the developer, to advance bona fide psychometric or policy purposes. The board did not drop the scores three SEM's to account for measurement error; the developer recommended a drop of only two SEM's for that purpose. Nor were scores dropped three SEM's to reduce adverse impact against blacks; the State Assistant Superintendent in charge of the certification test was vehemently opposed to taking race into account in setting the cut scores. Finally, while cut scores may have been lowered by three SEM's in part for the permissible purpose of maintaining an adequate teacher supply, the court is convinced that the primary purpose for dropping three SEM's was to mask the obvious system failure generated by the developer's cut-score methodology. For example, even after the developer's binomial twist, 78% of the teachers taking the first administration of core examinations would have failed, and the same would have been true for 93% of those taking the school counseling examination, 89% of those taking the learning disability examination, and 97% of those taking the library media examination. It is apparent that these pass rates did not reflect a fair construct of minimal competence. Further adjustments were employed to back into a passing rate that would appear tolerable and reasonable. The State Board of Education and the test developer in effect abandoned their cut-score methodology, with the result that arbitrariness, and not competence, became the touchstone for standard setting.

The court would be inclined to uphold the cut-score procedures employed for the nine examinations developed in 1982 and the ten examinations subjected to topicality review in 1983; however, each of these examinations has already been shown to be content invalid. Since a valid cut score cannot be generated by items that lack content validity, the validity of the cut-score procedure itself is not enough. Accordingly, the cut scores for the 1982 and 1983 examinations are also invalid.

In reaching the above conclusions, the court has been sensitive to a number of factors. First of all, as stated earlier, close scrutiny of any testing program of this magnitude will inevitably reveal numerous errors, and these errors will not be of equal footing. Secondly, cut scores cannot be determined with mathematical certainty, and political considerations may properly enter into cut-score decisions. The court's task therefore is to assess the sum gravity of the defects found, and to determine whether, as a result of these defects, the examinations are invalid as to content and cut scores. The court recognizes that, in carrying out this task, it must proceed with caution, and even deference. Although the court must assess the credibility of testimony advanced by each side and arrive at an independent judgment, the court should not readily set aside the findings of those who developed a test; the mere fact that the court sees things differently should not, by itself, be considered sufficient to impeach such findings. But while a court should eschew an idealistic view of test validity, it should also be careful not to apply an "anything *825 goes" view. In other words, the mere presence of conflict in expert testimony does not prove that a test fails to meet minimum standards; nor does it prove that a test meets such standards. A court should find a test invalid only if the evidence reflects that the test falls so far below acceptable and reasonable minimum standards that the test could not be reasonably understood to do what it purports to do. The court is convinced that this was the case with the Alabama Initial Teacher Certification test, and in particular with the early childhood education and elementary education examinations. [FN37]

FN37. The court recognizes that it has focussed not so much on the early childhood education and elementary education examinations, but on the Alabama Initial Teacher Certification Test as a whole. The court has done this because the history of the two examinations challenged by Richardson is the same as the history of the teacher certification test as a whole; the conclusions reached by the court regarding the certification test are also applicable to the two challenged examinations. Moreover, in order to appreciate fully the invalidity of the two challenged examinations, one must also understand just how bankrupt the overall methodology used by the

State Board and the test developer was.

The court also recognizes that it has focussed on the development and implementation of several individual examinations which have not been challenged by Richardson. The court has included these examinations as additional evidence of the invalidity of the State Board and test developer's overall methodology.

IV. RELIEF

Since Richardson is entitled to prevail on her disparate impact claim, the court must now determine her relief. The court will require that the Lamar County Board of Education reemploy Richardson as an elementary school teacher at a salary and with such employment benefits and job security as would normally accompany the position had she been employed in the school system since 1983. The court will also require that the school board pay her all backpay and other employment benefits she would have received had the school board reemployed her for the 1986-87 school year. The court will also require that the school board pay reasonable attorney's fees to her attorney. 42 U.S.C.A. § 2000e-5(k). The court will give Richardson and the school board an opportunity to agree, between themselves, to the appropriate amount of attorney's fees, present pay, backpay, and other employment benefits to which Richardson is entitled. If the parties are unable to agree, the court will then set these matters down for a hearing.

An appropriate judgment will be entered.

JUDGMENT AND INJUNCTION

In accordance with the memorandum opinion entered this date, it is the ORDER, JUDGMENT, and DECREE of the court:

1. That judgment be and it is hereby entered in favor of plaintiff Alice Richardson and against defendants Lamar County Board of Education and its superintendent and members;
2. That it be and it is hereby DECLARED that plaintiff Richardson may recover on her "disparate impact" claim but not on her "disparate treatment" claim against defendants Lamar County Board of Education and its superintendent and members;
3. That defendants Lamar County Board of Education and its superintendent and members, their officers, agents, servants, employees, attorneys, and those persons in active concert or participation with them who receive actual notice of this injunction by personal service or otherwise, be and they are each hereby ENJOINED and RESTRAINED from failing to reemploy forthwith plaintiff Richardson as an elementary school teacher in the Lamar County School System at a salary and with such employment benefits and job security as would normally accompany the position had she been employed in the school

system since 1983;

4. That plaintiff Richardson be and she is hereby awarded from defendants Lamar County Board of Education and its superintendent and members all backpay and other employment benefits she would have received had said defendants not illegally refused to reemploy her;
5. That plaintiff Richardson and defendants Lamar County Board of Education *826 and its superintendent and members be and they are hereby allowed 21 days from the date of this order to file a request for the court to determine the appropriate amount of present pay, backpay and other employment benefits to which plaintiff Richardson is entitled, should the parties be unable to agree to these matters;
6. That plaintiff Richardson be and she is hereby allowed 28 days from the date of this order to file a request for reasonable attorney's fees; and
7. That all other relief sought by plaintiff Richardson that is not specifically granted be and it is hereby denied.

It is further ORDERED that this court retains jurisdiction of this case until further order.

It is further ORDERED that all costs of these proceedings be and they are hereby taxed against defendants Lamar County Board of Education and its superintendent and members, for which execution may issue.

The clerk of the court is DIRECTED to issue a writ of injunction.

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

~ EPAA Vol. 7 No. 4: Haney, Fowler, Wheelock, Bebell & Malec "Massachusetts Teacher Test" ~

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

Appendix 3
Summary of Results of Interviews with Examinees

Our interviews with fifteen candidates who had sent us copies of their MTT score reports yielded background on the information available to test takers, administration of the test, including testing conditions, test content, test length, and test-takers attitudes about testing requirements for professional certification. In part 4 of this report, we summarized the nature of candidate views on these topics. Here we provide more detail on what was said by how many of our sample of 15.

Information available to test takers

According to teacher candidates interviewed, neither the testing company, National Evaluation Systems (NES), nor the Massachusetts Department of Education, provided adequate information about the test prior to administration. Regardless of whether they took the MTT in April, July or October, not one of the 15 teacher candidates interviewed reported receiving any useful information (or information they considered useful) from the Department of Education. Nine (60%) saw no study guide at any time. Two July test- takers called the Department of Education repeatedly and still failed to receive any information. Others reported receiving information that was misleading or inadequate, either regarding the consequences of the test or about test content or format.

Candidates reported that when they called the Massachusetts Department of Education, they were told no study guide was available. Of the candidates interviewed, all who contacted the DOE found that contact unsatisfactory. They described this contact and the information provided on the test as "not particularly helpful," "not readable," "vague" and "very limited."

Information for first-time test-takers

April test-takers registering for the test believed that test results would not "count" toward certification in Massachusetts. They reported that their colleges and universities also believed that because this was the first administration of the test, their scores would be used only to determine passing scores for future test-takers. This understanding was based on more than hearsay. The NES "1998-1999 Registration Bulletin" that many first-time candidates received reported:

> No qualifying score will be established until after the first

> two administrations of the tests on April 4 and July 11, 1998. Candidates who must take the tests and are eligible to participate in those first two administrations will satisfy the testing requirement by completing the tests. A qualifying score for each test will be determined by fall 1998 and used beginning with the October 3, 1998 administrations. From then on, candidates for provisional or provisional with advanced standing teacher certification who must take the tests will have to achieve a qualifying score to meet the certification requirement (p. 2).

One candidate prepared her own study guide based on Massachusetts curriculum frameworks, noted, "The six of us who graduated were not told much. Less than two weeks before the test date we were told we needed to receive a certain score to get a certificate."

July test-takers reported they received most of their information about the test from the media. Because of widespread media reports, some reported they were able to anticipate certain test components, specifically the dictation portion used to test "literacy and communication skills." In September 1998, the State Department of Education released a guide entitled "Massachusetts Teacher Tests: Test Information Booklet." The booklet includes test objectives, sample questions, and criteria for test scoring. This resource was unavailable to April and July test takers.

In the view of some candidates interviewed, the confusion about the test's consequences contributed to depressing test results. Whether they passed or failed, spring test- takers in particular attributed low test scores overall to the message that the test would not "count." One April test-taker noted, "They had told us the scores weren't going to matter. I thought it was a joke." Another reported:

> Many of us were in the middle of student teaching. We were really busy, and most people didn't study so much as I did. This could explain why so many people failed the first time around and then passed the next time.

A third candidate who passed all three portions of the test nonetheless remarked:

> I would guess that a number of people who took the very first test took it in panic mode. Because of all the hype, a lot of people went into the test in a panicked frame of mind. Under these circumstances, test scores are not an accurate representation of what they can do.

Information about Re-testing

Candidates interviewed also reported receiving inadequate information, if any, regarding a re-test for those who failed the test on the first round. For example, test- takers knew that free re-tests were offered to all candidates who failed one or more sections of the April test. However, one July test-taker reported that she did not know if such provisions would be made for those who failed the July test. Another candidate who failed the April test but did not intend to teach in Massachusetts described her uncertainty as to whether a re-test was still required because she had graduated from a private college in Massachusetts. She reported, "I was on the phone all the time. No one was giving me a definite answer."

Information about [illegible]st results

Several candidates observed that they received individual test results in a format that precluded them from learning from their mistakes. One who passed the test on the first round in April asserted, "It is not surprising at all that there are repeated re-tests. There is no information at all about what you did wrong, no information about what you did right." In part, these test-takers' wish for more precise information about their performance reflected some doubt about the objectivity of the questions and their scoring. As one noted:

> I'd like to see what I got wrong. A lot of questions were subjective and poorly worded. I could have supported any of my answers, even if they were marked incorrect.

Given the lack of either a study guide or individual feedback, some candidates were left with the belief that their results were a matter of chance. For example, one experienced teacher who passed still found the morning literacy portion of the test "very subjective." As she explained:

> For example, [on a question requiring the] summarizing of an article. I thought I did a good job, but my "bar" said I did an "adequate to inadequate" job on this. You could have three people grade this and come up with three different scores.

A July test-taker added:

> [There were] a couple of open-ended questions on how you would teach something. But different people could teach it in different ways. There was more than one right answer. There was no guidance, no rubrics. I would have liked to have seen a study guide.

Some candidates who passed the MTT on the second round attributed their scores as much to luck as to preparation. These candidates saw little difference between the knowledge they possessed at the time of their first test, compared with that knowledge at the time of their second test. Pressed to explain the differences, they alluded to differences in testing conditions, including more time to take specific sections of the test, familiarity with test format and content, or luck. For example, questioned as to what they thought had contributed to raising test scores from one administration to the next, one candidate who had failed the first round on the elementary portion of the test by two points replied, "It's a matter of luck." Another responded, "I don't know. Chance, totally chance!"

Testing Conditions

Candidates interviewed had taken the Massachusetts Teacher Test at seven different locations around Massachusetts. These testing sites included Bunker Hill Community College, Auburn High School, Burncote High School in Worcester, Malden High School, West Springfield High School, Randolph High School, and Wakefield Memorial High Schools. We asked the candidates to describe the testing conditions they experienced as they related to physical comfort, breaks allowed, adequacy of time to complete the test, acoustics, clarity of directions, and variations available for candidates with disabilities.

Most candidates generally found the general testing environment to be reasonable or on par with other tests taken. Problems at specific locations were noted by four candidates, including two July test-takers who mentioned heat and lack of air-conditioning as a problem, one who described proctors who talked "very loudly" among themselves for 20 minutes during the testing, and one who reported the test started late. However, almost half (47%) reported serious problems with acoustics that hindered performance on the written mechanics portion of the test. In addition, almost half also described problems related to the clarity of directions and test organization. Others raised concerns about the length of the test overall.

Acoustics

The morning portion of the MTT purports to test written mechanics by asking candidates to listen to a passage on a tape recorder, then write down the passage word for word. The purpose of this exercise is to "demonstrate the ability to spell, capitalize, and punctuate according to standards of edited American English" (Massachusetts Teacher Tests: Test Information Booklet, 1998, p. 9). Despite the apparently straightforward nature of the task, candidates reported a number of problems in its administration. Although eight candidates tested at four of the sites described the testing acoustics as "fine" or "very quiet," other candidates described conditions that hindered test performance. Five candidates (33.0%) tested at three different locations reported hearing tape recordings from other testing rooms. One test-taker voiced a common complaint:

> The acoustics were absolutely atrocious. The classrooms were adjacent to each other. I was near the back. I couldn't hear the tape recorder in my classroom, but I could hear the tape in the classroom behind me, and they were not in sync.

Clarity of test directions and format

In addition to variations in and problems associated with the quality of listening conditions, some candidates described problems with the clarity of test directions and test format. Specifically, while six candidates characterized the proctors' understanding and test's clarity of directions as "fine," "okay," or "not bad," nine others described proctors as "poorly prepared" or "kind of lost," and directions as "vague."

Several April test-takers noted that the lack of consistency in the test's format was a problem. As one explained, "You went from multiple choice to essay to fill- in-the-blank to correct a sentence that was incorrect." A July test-taker described confusing directions in relation to test items on the elementary education sub-test. For example, she explained that when she came to one test item, she was "not sure if they wanted a lesson plan or an essay." By the third administration, test organization and clarity of directions still raised concerns. As one October test-taker reported:

> I had a problem with [the clarity of directions]. They were extremely disorganized.... Because there were so many subtests...you had the blue section of Test A ...it was confusing, with too many subtests, colors, page numbers. It was terribly confusing. I don't see why they couldn't have had one test booklet. Also there was a misprint. They had misprinted the direction of where to answer something. They said it was on one page, but it was on another.

Two test-takers also described confusion related to the "tricky" selection of answers for multiple choice items. For example, one candidate estimated that up to 70% of the special education questions offered choices like "1 and 3 only," "2 and 1 only," "3 only," and "4 only."

Length of test

Whether they had passed or failed the first time they were tested, virtually all candidates mentioned the length of the test overall as excessive and believed it had an impact on their performance. One noted, "It was too long being in one room." Another reported, "Time was the big issue. Eight hours was just too long."

In particular, several candidates mentioned how the test's emphasis on writing essay questions or copying down dictated material contributed to increasing fatigue or incoherence over the course of the day. One candidate, an experienced teacher certified to teach high school English in New York, New Hampshire, and Connecticut and who described himself as a "good test-taker," compared the MTT to other state tests and reported:

> In other tests, a third was definitely related to education, a third was about literature, and a third had to do with analytical or problem solving skills. The New York test had only one essay. In Massachusetts, there was one in the morning, and two in the afternoon, plus the morning dictation. There was a lot of writing. I was used to it, but my handwriting's awful. By the time you get to the end of the day, you're not legible.

Another candidate, an experienced elementary teacher with teaching certification in two other states added:

> There was a lot of writing in terms of essays...summary of a story, essays on different topics, two more essays on elementary ed.... It was such a long process. Having the writing at the end was hard. In the end, I was exhausted. By the time I got to the end of the content test in elementary ed, I was just filling in the bubbles. My knowledge of the elementary subject area is much higher than I scored.

Second-time test takers faced some relief from the demands of an eight-hour testing day. Those taking only one or even two portions of the MTT as a re-test consistently reported that "it was easier" or "it was shorter," leaving them less tired. Candidates interviewed who took a re-test reported benefiting from the opportunity to re-rest under less pressure and with more time allowed. For example, one candidate who scored 60 on the writing portion of the test in April, passed with a score "in the high 80s" in July. Asked how she would account for her improved score, she reported:

> I knew I would be out in two hours because I was taking only the writing. I think the only thing I can think of is that I wasn't about to pass out. Maybe my handwriting was better because I only wrote for two hours.

> Another April-test taker who passed the writing portion of the test on the second try reported no difference either in her approach to the test or in the format or difficulty of the questions. However, she noted, "It was a lot shorter. I only had to take one part. Mine was the 8:00 to 10:00 slot, so I was done for the day, and I was not exhausted." Some repeat test-takers also noted that on the re-test, they had the entire morning to complete a section that was allotted only two hours the first time around.
>
> In commenting on the impact of the overall length of the MTT, candidates compared the MTT unfavorably to the length of other post-graduate tests. For example, one candidate who had also taken Graduate Record Examinations in her subject area noted that she had taken the GRE in four and half hours. Others noted that because it was possible to take the National Teachers Examination in sections, it was not so tiring. As one elementary teacher explained:
>
>> I took the pre-professional skills tests for the NTE on three separate days. You had different portions, but usually you'd come in for testing at 8:00 and be done by 10:30. NTE was offered every weekend, so I could break it up into different days.

Literacy and Communication Skills Content

Teacher candidates had a number of questions about the match between their "real world" literacy skills and test content. Whether they passed or failed the test, candidates reported that they found some of the content perplexing. Without any study guide or information from others who had taken the test before them, the expectations of April test-takers in particular were shaped primarily by other post-graduate tests. Given this, first-time test- takers were especially surprised at the dictation portion of the test, and several who had taken other teacher tests noted they had never encountered anything like this on any other examination. Generally, candidates were confused about what skills and knowledge the dictation portion of the test was attempting to assess. One successful October test-taker commented:

> The dictation section was mindless, a complete waste of time. They could have just given us the section and had us punctuate it. You don't have to write it down word for word. A monkey can do that.... It was very simple-minded.

Others also questioned other portions related to reading and writing. As one successful candidate reported:

> The reading and writing content didn't seem appropriate. [It

> was] not testing my knowledge of reading or writing. Like the question, "Define a verb." The test didn't seem to be testing what it said it was testing.

Subject Area Content

Candidates also detailed concerns about the content in the subject area tests. Their comments detailed concerns that much of the content did not match content found in the Massachusetts frameworks or the demands of real-world teaching responsibilities.

Content not included in grade-level curriculum frameworks

Three of the candidates, all of whom passed the test on the first round, believed the content tested extended beyond the knowledge they needed for teaching. One candidate mentioned that content included went well beyond what she would expect to teach her middle grades students. She asserted, "There should be separate tests for middle and high school science. Never will I teach advanced physics in middle school."

Another who took the English subject area test questioned the extent to which some content tested even fit within the boundaries of the discipline. As he commented:

> I distinctly remember a test question on "new journalism." I see this as a separate field. The question had to do with who was responsible for the rise of new journalism. I have a BA and an MA, and sometimes journalism students came into our classes, but we didn't take theirs. I've studied criticism, but not journalism.

Lack of content about professional knowledge

Some candidates interviewed reported surprise that the MTT did not cover content they expected based on their experience taking other teacher tests, whether the National Teachers Examination (Praxis) or state tests. Most often, candidates mentioned the lack of content related to professional knowledge and skills and commented on the lack of attention to teaching itself. For example, one recent graduate and physical education teacher noted:

> The content wasn't fair. There was not a lot of application to how you taught. The test did not have, "In this situation, what would you do?" kinds of questions.

An experienced elementary teacher certified in two fields in another state likewise wondered how well the MTT could determine teachers' competence in classrooms given the mismatch she perceived between the demands of teaching and content on the MTT. She asserted:

> The MTT did not test my ability to teach. As a teacher, I'm constantly reteaching myself. There are so many things you do as teacher that are performance- based, not knowledge-based. You need to know teachers are performing, handling their class in a professional manner.

Another teacher with twenty years experience and certification in two other states, also questioned what she perceived as an overall stress on mechanical skills and content knowledge compared to the limited emphasis on knowledge of classroom practice. She reported:

> There was a big lack of anything to do with classroom management. It was *all* content knowledge. There were no more than a couple of questions on teaching methods, classroom management, anything to do with your ability to teach.

In light of limited content geared toward assessing professional knowledge, candidates raised concerns that the emphasis on content to the exclusion of other aspects of teaching reflected a misunderstanding of what was required to be a successful teacher. As one candidate with an exemplary academic record asserted:

> Nothing on the whole test addressed the issue of how well you teach. Even if I'd gotten every question right, it wouldn't have proved I could teach physics to 16-year-olds.

From the perspective of another candidate, even when test items ostensibly melded content knowledge with teaching knowledge, the questions did not provide adequate information to answer them well. For example, one experienced teacher described one question on the middle school portion of the test as unrelated to real-world teaching conditions:

> You were supposed to create a unit plan with a team, but in the test, you're on your own. Then you're told you're not being graded on the creativity or usability of your plan. But there are no references, resources, or curriculum frameworks to work with. You're told your response will be graded on your knowledge of your subject area, not on the lesson plan.

Candidates viewed the weighting of basic literacy and writing skills over teaching knowledge, regardless of the subject area, as a major flaw in the test. As one explained:

> It's possible with the Massachusetts test for some very good teachers to be knocked out because of problems with spelling. A lot of what we do [in classrooms] is done ahead of time. It's not a handicapping situation not to know

> everything there is to know. It's also possible that a lot of bad teachers are passing. People could pass the content test without knowing how to teach.

Views of teacher testing

Despite the range of concerns raised about the MTT, the candidates interviewed did not object to standardized testing per se, and all agreed it was reasonable to ask teacher candidates to pass a test prior to certification. However, they added that although they were not opposed to the testing of new teachers, they believed that the MTT should be replaced with a different test. Whether they passed or failed, candidates interviewed were aware that "most of the professions have a test" and viewed testing one of the rites of passage into a profession. However, candidates doubted that the Massachusetts test in particular could adequately assess teacher competence. They compared the MTT unfavorably to other professional tests they had taken and described the latter as having a more balanced focus on all aspects of teaching, including content knowledge, classroom skills, and problem solving. As one noted, "Yes, I'm in favor of a fair test like those offered in Connecticut or New York. But this one is not fair at all."

Candidates noted in particular that, compared with other tests, items pertaining to the teaching process itself were missing from the Massachusetts test. As one explained:

> In New York, you have to pass a basic liberal arts and sciences test.... But in New York you have to pass a written assessment of teaching skills. This includes subject knowledge but also more teaching. There are also two essays associated with classroom situations.

A third elaborated:

> It's not unreasonable to have new teachers take a standardized test. It's only unreasonable to take an unproven test that has no validation data. This one is much too subjective. The New York test was more fair in assessing teachers' ability to teach.... If they want to know if you can write under pressure, that's what this test shows.

Eight candidates explicitly mentioned the National Teacher Examination as a more accurate test of teaching skills. One experienced teacher who taught elementary school in two other states before moving to Massachusetts asserted, "The Massachusetts Teacher Test needs to look into the NTE. It's recognized, accepted, reliable." Another described the NTE as testing "a wider range of content--communications, professional knowledge, and general knowledge." Another concluded, "Massachusetts could do better, could have a much better quality test."

page 1 | introduction | background | reliability & validity | interviews | conclusions | references

• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 279 times since February 17, 1999

Education Policy Analysis Archives

Volume 7 Number 5 February 17, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ
85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Some Comments on the Ad Hoc Committee's Critique of the Massachusetts Teacher Tests

Howard Wainer
Educational Testing Service

> "It is a trite but true observation, that examples work more forcibly on the mind than precepts: And if this be just in what is odious and blameable, it is more strongly so in what is amiable and praise-worthy."
>
> Henry Fielding, *Joseph Andrews*, 1742

The critique of the Massachusetts Teacher Tests (MTT) by Haney and his colleagues is deserving of comment, both because of the impact of the MTT and because of the evocative manner in which the tale is told. Their emphasis on examples makes for a forceful argument, and I fear that my reliance on precepts may look meager by comparison. Nevertheless, I hope that some of the observations that follow contribute to the more reasoned assessment of these instruments and their use not just in Massachusetts but in the many other states where similar programs are being developed or contemplated.

It seems clear from some of the data that the Ad Hoc Committee have presented that the MTT is not up to snuff, although it is a bit of a puzzle why its reliability isn't higher. Even a little bit of pre-testing (and Spearman- Brown) would have shown what test length is required for standard reliability. Perhaps time and economic pressures intervened and a less than fully developed instrument was rushed to the field? As the old saying goes, if it's worth doing, it's worth doing badly.

The reliability in this whole process that is of greatest interest is the reliability associated with the pass/fail decision (i.e., the reliability that emanates from the standard error of measurement around a score of 70). This can be easily calculated from the raw data by noting the inverse of the information function after fitting an Item Response Theory (IRT) model. If the raw data were available, I could do it myself. Considering the high stakes associated with these tests, there is more than the usual obligation on someone's part to make these raw data available.

It can be noted in Figure 1 that 32 of the 66 teachers who failed initially passed on retesting. This does not necessarily speak to unreliability of the tests. It is possible that grit, determination and plenty of quick preparation accounted for the elevation of the scores on the second administration. One wishes to see the other half of the four-fold table (the counts of those who passed the first time and who passed or failed the second time), but such data are never available--passing once is a ticket out of the testing system. One could contrive an approximation (in the spirit of split-halves) by generating two scores (e.g., odd and even item scores) and constructing the 2-by-2 table (pass/fail vs. score-1/score- 2) using 35 as the passing score. Of course, one would want to do this for all candidates, not just those who failed the first time. The extent to which this table is diagonal (i.e., approaching zeroes in the off-diagonals) is the reliability of the test for the decision. Of course it is an underestimate (too short a test) and you need to apply something like the Spearman-Brown "prophecy formula" to estimate the reliability of the decision based on the full test.

A comment made about "Peter" in the examinee vignettes seemed a bit strong. Peter scored in the 91st percentile on the GRE-V and "between the 80th and 85th percentile in Reading". These two results were described as "quite at odds". In fact, they seem fairly consistent, especially considering the somewhat lower reliability of the MTT.

One might even say that the Ad Hoc Committee exaggerated the discrepancy between Peter's SAT and MTT scores for effect. (Indeed, a more temperate, less polemical, tone throughout the article would have been more persuasive for me and most readers, perhaps.) This same bit of exaggeration shows itself again in Recommendation 1: "No exam at all is better than an unreliable exam...." Whether or not no exam is better than any exam depends on (i) how unreliable that exam is, (ii) what the selection ratio is, and (iii) the nature of the cost function associated with errors of each type.

For example, suppose one has a test that is 92% accurate (4% that should "pass" actually fail and 4% of those that should "fail" actually pass), and suppose further that one has a selection ratio of 95 to 100--that is one expects to pass 95% of all applicants, and false passes are as bad as false failures. Under these circumstances, the test is a bad idea since if one simply passed everyone the error rate would be 62.5% of what would be obtained if the test were used. But suppose the cost function is different. Suppose a person who fails improperly can take the test again and pass whereas a person who passes improperly is installed forever to do irreparable damage to our children. Now, using the test with its obvious imperfections seems like a better idea.

What cost structure would one wish to impose on a licensing exam for heart surgeons? Airline pilots?

And what about validity? The ancient Chinese tests had an enormous selection ratio--one in a "gazillion." Consequently, the test did not need much validity to be worthwhile, though it did need at least some. Of course enormous numbers of worthy candidates failed, but the odds of an unworthy one passing were small enough to be ignored for all practical purposes. The same structure manifests itself in such tests as the National Merit Scholarship test in which 1,500 "winners" are chosen from more than 1,500,000 applicants. This 1 in a 1,000 selection ratio means that a very large number of worthy kids do not win, but that all winners are truly wondrous. These three key issues that must be decided before making such a statement as "no exam at all is better than an unreliable exam" where "unreliable" means "reliability in the 0.7 range".

I conclude, then, on the basis of my reading and my own biases, that the MTT needs work, but certainly provides some information to guide decisions. And, if the selection ratio is high enough, the MTT might even be good enough. (I tend to put a high cost on allowing an incompetent teacher into the classroom and a relatively low cost on asking a marginally passing teacher to take the test again.) But that's just me.

I hope that the Massachusetts Department of Education will now make enough data available for a more proper test analysis and, perhaps, publishing the Ad Hoc Committee's critique will hasten that day.

About the Author

Howard Wainer

Email: hwainer@ets.org

Howard Wainer received his Ph.D. from Princeton University in 1968, after which he was on the faculty of the University of Chicago. He worked at the Bureau of Social Science Research in Washington during the Carter Administration, and is now Principal Research Scientist at the Educational Testing Service. He was awarded the Educational Testing Service's Senior Scientist Award in 1990 and was selected for the Lady Davis Prize. He is a Fellow of the American Statistical Association. His latest book, *Visual Revelations*, was published by Copernicus Books (a division of Springer-Verlag) in 1997.

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North
Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 444 times since February 18, 1999

Education Policy Analysis Archives

Volume 7 Number 6 February 18, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

College Students' Use of the Internet

Anna C. McFadden
University of Alabama

Introduction

Over the last several years there has been mounting concern about children being exposed to sex-related material on the Internet. Concern about pornography and obscenity is widespread and this concern has spawned a host of products to block or filter content. The notorious *Time* magazine article (July 3, 1995) "Cyberporn"--which *Time* later acknowledged had doubtful credibility (July 24, 1995)--undoubtedly inflamed this trend. The article, which asserted that much of traffic on the Internet dealt with pornography, was based on the largely discredited research of a Carnegie Mellon undergraduate student who examined 32 alt.binaries newsgroups on Usenet, not the Internet. Nonetheless, the article was fodder for the Communications Decency Act of 1996. While the Supreme Court struck down the Act, pending bills such as the "Safe Schools Internet Act" (H.R. 3177) would require all public libraries and schools that receive federal funds for Internet access to install blocking software to restrict minors' access to "inappropriate" material. Other pending bills would punish commercial online distributors for access to material they do not directly control and require service providers to offer blocking software to customers.

While most students who use computers in university computer labs are legally adults, many are not. If laws restrict access to minors, there will be a host of technical problems to provide access to scholars and adult students. Labs are open spaces where students come and go, using computers for many purposes but only part of the time for Internet access. Determining policies and creating procedures to implement and monitor policies will entail considerable resources for something that may not be a serious problem and something that cannot be effectively controlled with filtering software. It could require students to present identification to prove they are adults in order to access certain computer resources, not to mention the procedures that would be used to restrict access to those who are minors. There is no way to verify age on the Internet, so the responsibility would fall to the school staff. For the time being, most universities have policies that limit computer use to legitimate educational purposes, and students in most universities have mainly unrestricted access. There is little or no information about how the Internet is used in such settings.

Purpose

The purpose of this study was to determine the nature of Internet uses by students in a computer lab of a major state university. Of particular interest was the percentage of "hits" associated with pornography and gambling sites.

Setting

The study was conducted in an "open" computer lab of a major state university. The lab contained 62 computers, 2 of which were inoperative at the time of the study. The lab is available to any student on the campus and is open six days a week for approximately 70 hours each week. Any student may simply enter the lab, take a seat, and begin using the computer. All 60 computers had Internet connections, ear phones, CD- ROM, and were equipped with standard tools, such as word processing, database, spreadsheet, presentation software, graphics, statistics software, and other applications.

Method

Due to the fact that the computers in the lab are open, no computer is assigned to any particular student or group. There was no way or any wish to identify any student who used a computer, and all computers have several different users each day. Also, students do not store personal files on the computer but keep their files either on a literal drive on the network or copy their files to floppy or zip disks. A student may use any particular application for long periods of time, such as word processing or a statistical package, so Internet use is intermittent.

The researchers elected to randomly select 10 percent of the computers for the study. Thus, 6 computers were selected with the aid of a table of random numbers. The researchers copied the Internet cache of the six computers at the same time in the late afternoon on January 25, 1999. This date was selected, in part, because it was assumed that dates close to holidays (e.g., Christmas, Halloween, and Valentine's Day) would lead to spurious results with "hits" unrelated to normal usage. It should be emphasized that the cache provides anonymous information in an open lab. No attempt was made to examine the cache of any office computer because this would be an intrusion on privacy, and only computers in the open lab were studied. No attempt was made to determine how long each computer was used for Internet activity or the amount of time in any particular category of Internet activity.

Results and Discussion

The six computers had a total of 2,310 Internet "hits" stored in cache for an average of 385 per computer. This seemed to be rather small, perhaps because the cache limit was set at low levels for two reasons: (a) different users do not benefit from cache memory as a single user would, and (b) a lower cache limit frees up space on computers that are heavily loaded with many different software programs. Nonetheless, the cache records served as a useful sample of activity for purposes of this study.

It should be emphasized that the percentage of "hits" indicates the sites contacted, not the percentage of time the computer was actually used for any purpose. A computer may be off line for many hours for word processing or the amount of time a student accesses a "course site" may be for the purposes of getting an assignment. The data were organized into convenient categories for data analysis, although our intent was to examine the contents specifically for sexually explicit or gambling related sites. Nearly half of Internet use was accounted for by a category labeled "General Sites." Due to the number and diversity of these sites, it was decided to categorize them under this general heading. These included sites apparently related to course activities, research, or personal interest, including anatomy, science, books, literature, airlines, government web sites, health and disease, psychology, business statistics, and the like. This is the largest category reported because of the way the data were collapsed for categorization. The complete categorization of "hits" appears in Table 1 below.

Table 1
Number and Percentage of Internet Use

Type of Site	Count	Percent
Sites (General)	1094	47%
Mail	647	28%
Chat	133	6%
Search	133	6%
Sports	137	6%
Course Sites	102	4%
News	30	1%
Sex	29	1%
Games	2	0%
Radio	3	0%
Total	2310	99%

(Note: Due to rounding the total does not reach 100% and games and net radio use are below 1%.)

No gambling sites (0%) were contacted and contact with sexually explicit sites was low (1%). Many e-mails were sent or received (28%). Course-related activities as well as personal use might account for much of this e-mail, although personal use is probably high. Internet Chat represented 6% of activity, some of which may also be accounted for by course requirements. An equal percentage of activity involved contacting sports related sites (6%), such as ESPN, university sports pages, and the like.

If this study is representative of the college population, the overwhelming use of the Internet in an open computer lab conforms to university acceptable use policy. There is far less use of lab computers to contact pornographic sites than we have been led to believe; in fact, there was almost none. There was as much interest in news (1%) from online sources, such as MSN or CNN, as there was interest in sex. Pending legislation that could impact university labs would require a potentially expensive and cumbersome procedure for a very minor problem. This "snapshot" of college student computer use on the Internet reveals a picture remarkablke in its "ordinariness."

About the Author

Anna C. McFadden

Associate Professor
Instructional Technology
The University of Alabama
Tuscaloosa, Alabama

Phone: 205-333-9185 Fax: 205-333-8288

Homepage: www.bamaed.ua.edu/it

Dr. McFadden teaches doctoral-level courses in instructional technology, including local area network management and development of online media. She is leading a team in initiating Real Audio/Real Video development as part of the web-assisted classes, focusing on asynchronous learning activities. Dr. McFadden and her team, over the last twenty years together, have secured in excess of $6 million in external funding for research and development activities. She also serves as the Editor for Internet In-Sites , a monthly online newsletter for instructional technologists in K-12 schools around the world. She also serves on the Editorial Board for The Internet Source , a quarterly journal for teachers in international schools. As a senior partner in emTech Consulting www.emTech.net, Dr. McFadden is involved in professional development activities in K-12 schools in the USA and around the world.

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa- DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 214 times since March 1, 1999

Education Policy Analysis Archives

Volume 7 Number 7 March 1, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287- 2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Restructuring the Schoolbook Provision System in Indonesia: Some Recent Initiatives

Dedi Supriadi
Institute of Teaching and Educational Sciences
Bandung-Indonesia

Abstract

Schoolbooks account for between 65% and 75% of all publishing activity in Indonesia. The amount of money allocated is continually increasing. Priority is given to the primary and junior secondary school levels (6+3 years), which are compulsory. Between 1969 and 1988, the Government of Indonesia (GOI) has produced some 550 million primary textbooks and library books. Up to the year 2000, the GOI has decided to allocate the total of US$ 355.2 million to produce 250 million copies of primary and junior secondary school textbooks to reach the ratio of one book for each student. In addition, around Rp 20-50 billion (US$ 10-20 million) annually is spent to purchase 8-17 million copies of reading books which are aimed at stimulating the reading interest of primary school children. These books are distributed free to some 168,000 primary and 26,969 junior secondary schools throughout Indonesia. Following the massive efforts to increase book availability at schools, some innovative policies are being taken. Book evaluation standards have been improved to ensure that only high quality books are used at schools. The distribution system has been restructured to guarantee that books reach targeted schools. Consequently, the book monitoring system has had to be strengthened to examine whether or not the books really reach the schools and are used properly by students and teachers in the classrooms. In the last three years, there has also been a growing concern with multicultural issues in schoolbook provision programs. In such a culturally diverse nation as Indonesia, schoolbooks should also be culturally sensitive and be recognize the varied sociocultural backgrounds which affect students' learning.

Schoolbooks and Educational Quality: International Comparison

Schoolbooks, especially textbooks, tend to be the dominant instructional medium in the classroom (Patrick, 1988). Because textbooks typically deliver the curriculum, they are regarded as the single most important instructional material. As a result, they are central to schooling at all levels (Lockeed & Verspoor, 1990). Especially in a context of scarce learning resources and a shortage of teachers--in terms of number and quality-- textbooks appear to have a strong positive impact on educational quality, as indicated by student achievement. A review by Heyneman, Farrell, and Salveda-Stuarto (1981) has found that of 18 correlational studies, 83% showed statistically significant positive correlations between textbook availability and student achievement.

In the Philippines, increasing the student-textbook ratio from 10:1 to 2:1 in the first and second grade of primary schools in the late 1970s led to significant learning gains among students using the new textbooks. In Nicaragua, students who received textbooks scored significantly higher by about one-third of a standard deviation on a test of mathematics achievement than students not supplied with textbooks. In Brazil, the compensatory program of adding various "school quality" elements--especially textbooks--to poor rural schools in the 1980s confirmed the crucial importance of textbooks; students made impressive improvement in achievement on tests of mathematics (World Bank, 1995: 10-11). In Indonesia, a national survey in the 1970s found that students who had library books and many other classroom facilities, as well as good teachers, demonstrated higher achievement (BP3K, 1980).

Realizing these facts, many developing countries invest heavily in schoolbooks, especially textbooks, with higher priority given to primary and secondary school levels. Many developing countries are now moving toward pursuing the goal of "a textbook for each" student. But countries pursue this goal in different ways. One can ask of these strategies: who are the targets, what is the book distribution system, what are the roles of government and private sector, and what is the contribution of the students' parents? Ten countries assisted in this endeavor by World Bank loans can be cited as examples, as summarized in Table 1.

Table 1
World Bank/IBRD Textbook Assisted Projects in Ten Countries

Country	Targeted Areas	Number of Textbooks	Distribution scheme
Indonesia entire country	• 29.2 million primary school students procured by MOEC's Book Center	158 million copies of student books+) & 6.3 million copies of teacher guides 100% funded by GOI.	Free for all students Total cost: US $ 182.2 million.
	• 8.8 million junior secondary school students, procured by private publishers	85.1 million copies of student books++) & 980 thousand copies of teacher guides, cost: US$ 172.8 million.	Free for all students
India Utar Pradesh	Primary schools, procured by private publishers.	83 million copies, reaches all children	Parents purchase at lower prices
China entire country	Primary schools, produced by state	2.2 billion copies for 180 million students (!)	Parents purchase at subsidized prices
Vietnam entire country	Primary schools, procured by state	Textbooks for all children, Per book prices: US$ .15	Parents purchase, allowing choice of buying or renting
Brazil, Northeast Provinces	Primary schools, produced by private publishers	100 million copies per year, Total cost: US$ 63.1 million	Total cost: US$ 63.1 million under 1/1 basis
Romania, entire country	Primary schools, produced by private publishers	15.7 million copies, total cost: US$ 41.4 million	Textbooks are provided to schools for purchasing
Chile, entire country	Primary schools, produced by private publishers	Total costs: US$ 25 million	Free for students in deprived urban and rural areas

Philipines, entire country	Primary schools, produced by government	44 million books, total cost: US$ 50 million	Free books for the poor, rich parents could purchase
Venezuela, entire country	Basic education, procured by private sector	11.3 million copies, total cost: US$ 23 million	Free for all students in marginal areas
Mexico, four states, rural areas	Basic education	70 million copies per year, many books never reach poor provinces	Free for all students in all schools in project

+) Covering 6 subjects Pancasila, Bahasa Indonesia, Mathematics, Science, Social Studies, and Local content

++) For 9 subjects Bahasa Indonesia, Mathematics, Physics, Biology, History, Economics, Geography, English, Art and Craft

Source World Bank, 1995 *Staff Appraisal Report Indonesia. Book and Reading Development Project Report No 141016-IND*, pp 21, 32

The Anatomy of Schoolbooks in Indonesia: Vary in Kinds, Complex Procedures of Provision

The nature of schoolbooks in Indonesia is quite complex, both in terms of kinds and procedures of provision. Schoolbooks consist of student textbooks, teacher guide books, reading books, and reference books for teachers and students. There are two kinds of textbooks: 1) government textbooks which are obligatory for school use, and 2) supplementary textbooks which are not obligatory, depending on the needs of each school. Government textbooks used to be published by Government (Ministry of Education and Culture, MOEC) mainly under the coordination of the Book Center and other projects at the Directorate General of Primary and Secondary Education (DPSPE). Supplementary textbooks, on the other hand, are published by private publishers. Teacher guide books used to be provided together with student textbooks.

Reading books which consist of fiction and nonfiction books are those which are intended to promote reading interests and increase literacy of primary school children. Thus far, the effort to provide reading books is only for primary school level, while for the upper levels (junior and senior secondary) priority is given to textbooks (production, quality, and distribution). Annually, each of the 168,000 SDs/MIs (Note 1) throughout the country used to receive between 50 and 100 copies of reading books in different titles. Counting all primary school libraries, approximately 75 million reading books have been delivered to schools since 1979.

The main issue concerning these books is how to maximize their use by pupils so that they can have significant impact on pupils' reading behaviors. Low correlations between the presence of reading books and student reading behaviors as revealed in the 1989 IEA (the International Association for the Evaluation of Educational Achievement) study (Note 2)and cited in the 1992 World Bank's report do not necessarily mean that those books are not useful, rather they indicate that the use of these books by students is still very low. This was supported by the fact that many students spend relatively little time "reading for fun." At grade four, approximately one in five (19.5%) claimed to read a book for fun "almost never." Almost 60% of students considered themselves "average" or "not very good" at reading, while only 27.8% considered themselves "good" and 14.4% "very good." It is for that reason the DGPSE continues to work with students (and teachers) to maximize the use of reading books at primary school libraries.

Book Evaluation and Approvals: Procedures and Criteria

To be eligible to be used at schools, both supplementary textbooks and reading books have to pass the evaluation process carried out by the National Evaluation Committee for private publishing books (which I have served as chairman since 1993) and be officially approved by the Director General of PSE. The Committee members consist of experts from universities, IKIPs, and other government agencies.

The evaluation comprises four aspects, viz., quality of content, language, graphical presentation, and security.

Quality of content refers to such indicators as scientific truthfulness and its relevance to the existing National Curriculum. Language basically means readability, legibility, and consistency with the standardized Indonesian language. Graphical presentation includes quality of printing, binding, format, illustrations, colors, typefaces, font sizes, spacing and book appearance in general. Security means that the content of the books should be in line with and not contradictory to Pancasila (the Five Principles--Note 3), UUD 1945 (National Constitution), Government policies, national unity and security, laws, regulations, ethics, and that the content not exploit the sensitive issue of SARA (ethnics, religions, race, and intergroup relations). Judgment of this aspect is made by evaluators coming from Mabes ABRI (Armed Forces Headquarter), Kejaksaan Agung (Office of Attorney General), Lemhanas (National Defense Institute), and Inspectorate General of MOEC.

The number of books considered in the evaluation process has increased over the years, indicating the rise of market demand, the productivity of book writers, and growth in the size of the schoolbook sector. In 1995, the number of books nearly doubled over previous years. However, the quality of books has not increased significantly as shown in the percentage of books passed in the evaluation (Table 2). To bridge this deficit, the Directorate of Educational Facilities, DGPSE, has repeatedly held meetings with book writers, editors, and publishers to address some basic mistakes found in the evaluation process. The DGPSE has also suggested ways of dealing with those mistakes to publishers.

At the same time, the Book Center under the Secretary General of MOEC has been conducting workshops and training sessions involving book writers and editors. Some of the participants in these activities are newcomers but appear to have potential for book writing. This program could help alleviate the shortages and the low quality of many books.

These efforts are expected to have significant effects on the capacity of writers and editors to produce high quality books in response to the rising demand for higher quality schoolbooks. The criterion of evaluation for supplementary books has been raised from a minimum of 60 to 80, to ensure that only high quality books are recommended for use in schools. It is particularly important for supplementary textbooks to be qualitatively comparable to the main textbooks evaluated by the National Textbook Evaluation Committee (NTEC). Another criterion is that supplementary textbooks must contain significant amounts of enrichment material.

Table 2
The Number of schoolbooks participating in the evaluation process at the Directorate of Educational Facilities, DGSPE - MOEC, 1979-1995

Year	No. of Books	No. Passed	%	No. of Publishers
1979	2,860	1,078	38%	n.a
1980	1,287	747	58%	n.a
1981	1,847	808	44%	n.a
1982	1,432	956	67%	n.a
1983	1,786	1,257	70%	n.a
1984	2,551	1,099	43%	n.a
1985	3,645	1,812	50%	n.a
1986	2,536	1,232	49%	n.a
1987	1,132	567	50%	n.a
1988	2,079	989	48%	n.a
1989	2,740	1,134	41%	n.a
1990	2,763	1,570	57%	101
1991	2,179	781	36%	126
1992	2,685	1,455	54%	170
1993	2,565	1,359	53%	237
1994	2,380	848	36%	230
1995	4,162	1,437	35%	275
Total	40,629	19,462	48%	

Source Directorate of Educational Facilities, DGPSE-MOEC, 1995

Junior Secondary School Textbooks: Improving Textbook Quality, Restructuring Distribution System

For the junior secondary school level, the GOI has launched a new project--effective of October 1, 1995-April 1, 2000-- called the Book and Reading Development Project (BRDP). The Project aims to improve the quality of education for secondary school students (nine-year compulsory basic education). In particular, the Project is expected to contribute to improving production and distribution of more and better quality textbooks, their availability in the classrooms, and their use by teachers. These objectives are to be achieved by assisting in the implementation of the textbook provision and distribution system to all school students on a sustainable basis, giving help to teachers in how to make better use of the books in the classroom, and motivate and promote reading habits among children (World Bank, 1995a).

Different from the previous book projects, the BRDP is developed under some innovative policies. First, books are selected by the NTEC--chaired by the Director General of PSE-- based on a set of high standards and strict evaluation procedures to guarantee that only high quality textbooks will pass the evaluation. Members of the Committee represent those experts whose achievement, commitment, and personal integrity have been widely acknowledged.

Secondly, books are selected from private publishers and should consist of both student books and teacher guides for grade one to three. Thirdly, book selection and procurement are "liberalized" in that international publishers--including those who are known as "global players" in the publishing business-- could participate in a bidding procedure known as International Competitive Bidding. Fourth, production and distribution of books are given to publishers whose books are selected to be purchased. Fifth, books are directly delivered from the publishers' warehouses to each individual school, rather than following conventional routes as from Kanwil's (Provincial Office of MOEC) warehouses down to Kandep (District Office of MOEC) and then sent to schools. Sixth, textbooks are procured at the provincial level --in which 90% of the total expenditures are allocated--and for that, each Kanwil is given authority to select only one title for each subject and grade for distribution to all schools in a given province.

These innovative policies have significantly changed the entire scheme of book provision with government tending to become the regulator and the private sector the main actors--a trend that is also occurring in many other sectors.

Implication of Reform: The Crucial Importance of Strengthening Monitoring and Evaluation

Reforms in the book distribution system are aimed at strengthening monitoring and evaluation procedures to ensure that books reach the schools and are used in the teaching-learning process so that they will have significant impact on quality improvement. (Note 4) The monitoring activities are focused on providing timely and accurate information regarding actual receipt of books in the schools and their use. Three main indicators are used: (1) how well teachers, parents, and administrators understand the objectives and procedures of the system?; (2) how effectively books are distributed to schools?; and (3) how effectively teachers and students use them in the classroom, and what is the effect of the books on quality improvement?

These activities involve 29 state IKIPs and Schools of Education at Universities in 27 provinces, and the Office of Inspectorate General of MOEC. Since 1996, two kinds of activities have been carried out: conducting an indepth study of 162 primary schools (6 schools in each province); conducting a quantitative survey of 810 primary and 810 junior secondary schools (30 schools in each province) to get baseline data on book condition at schools.

Each shipment of books to a school includes a simple self- addressed, postage-paid postcard which provides the following information: (a) date books received; (b) condition upon arrival; and (c) sufficient quantity or not. The returned postcards are then analyzed and reported by independent experts of IKIP. More in-depth and comprehensive studies are conducted by independent experts of IKIP and the universities. They employ both quantitative and qualitative methods to gather data/information through a series of interviews, observations, documentary studies, and checklists.

During school visits which take 5 to 6 days in each sampled school, they conduct interviews with principals, teachers, and parents, asking specific questions about the new textbooks, how well they are liked, if they seem to be helpful, problems related to book deliveries, etc. They also interview a small number of students asking similar questions, and observe and report activities of two or three classroom sessions involving textbooks. Sampled school are assigned to represent different locations, i.e. urban, suburban, and rural; public and private, Java and outside Java, better served and under-served schools.

Reading Promotion Program: In Response to the Challenges of 21st Century

The overall on-going efforts to increase book availability at schools (to reach the ratio of 1:1--one book for each student at basic education) and to improve their quality are also followed by a reading promotion program. This is a priority in Indonesia at present and has officially been declared as a national movement by President Soeharto on May 1995. This massive program is based on a belief that the creation of a reading society is a must, a conditio sine qua non, for the nation to successfully enter the 21st Century. The month of May every year has been declared "The National Book Month" and September has been declared "The Month of Visiting Library." In December 1995, the Society Association for Reading Interest Promotion (Perhimpunan Masyarakat Gemar Membaca, PMGM) was established in Jakarta representing those individual citizens concerned with reading promotion.

In addition, at the Book and Reading Development Project, there is a component dealing with reading promotion under the coordination of MOEC's Center for Communication Technology (Pustekom) in cooperation with PMGM. Its main targets are primary and secondary school students in addition to the community at large.

The challenge of book and reading promotion in Indonesia is obviously not easy, since many facts show that reading interest and behavior of Indonesian people, including school children, are still low. For instance, IEA has conducted a study covering the samples of 3,169 fourth graders of primary (SD) and 1,929 second graders of junior secondary schools (SLTP) from seven selected provinces. The sampled provinces of SDs were Riau, Jakarta, West Java, Central Java, Yogyakarta, East Java, and East Nusa Tenggara; while for SLTPs were Riau, Lampung, Jakarta, West Java, Yogyakarta, East Java, and East Nusa Tenggara.

One of the most disturbing findings is that Indonesian primary school students placed 26th out of 27 countries participating in the study, mastering only 36% of test items. Of the 27 countries, the highest scores were recorded by Finish students and the lowest were by Venezuelans. In Indonesia, among the seven provinces, the highest average scores were achieved by Yogyakarta (49%), followed by Jakarta (44.5%), East Java (38.6%), Riau (36.5%), Central Java (34.8%), East Nusa Tenggara (32.6%), and West Java (32%). At the junior secondary school level, Indonesian students placed 5th of 5 countries surveyed in Asia, i.e. Hong Kong 75.5%, Singapore 74%, Thailand 65.1%, the Philippines 52.6%, and Indonesia 51.7% (World Bank, 1992: 15-16; World Bank, 1994: 74).

This research has also revealed that both at primary and junior secondary schools, home-related variables have a greater effect on student reading behavior than school or teacher-related factors. This finding points toward the importance of massive involvement of families and parents in any effort at reading promotion.

Book Provision for Multicultural Students: Some Recent Development

Another development is that in the last three years there has been a growing concern with the importance of considering multicultural issues into schoolbook provision programs. This concern grows out of is a belief that in such a culturally diverse country as Indonesia, schoolbooks should also be culturally sensitive and recognize varied students' sociocultural backgrounds which affect their learning.

How is cultural diversity taken into account in schoolbook provision in Indonesia? In 1993, the DGPSE initiated a substantial effort to recognize student sociocultural backgrounds in reading book selection and provision. Preceding the decision taken by the Director General of which books to be purchased by the project, books are read by sampled students in different regions. The goal is to identify which books are more interesting to children. Each title is read by 10 students from different locations, and samples consist of fourth, fifth, and sixth graders of primary schools. Three indicators are used, i.e.: degree of "interestingness," readability, and meaningfulness. Annually, no fewer than 600 titles are read by students, of which some 400 titles meet the criteria. On the basis of this survey, the DGPSE classifies the books into two categories: 1) those books assumed to be "culturally-fair" are sent to all or most of schools; 2) those books which tend to be "culturally-bound" are allocated to schools in certain provinces-- usually around 10-15% of the total book titles. However some locally-content reading books are also assigned to the national level (distributed to schools outside its own province) to enrich students with knowledge of other cultures.

In spite of many commonalities in choice, it was evident that sociocultural backgrounds played an important role in affecting student choices. In a given geographical location, there were books which were highly favored by students (as indicated by scores) while in other locations the same books scored low. In more specific cases, such books as the folk stories of Wayan Bunter of Bali, Si Kabayan of Priangan- West Java, Si Malin Kundang of Minangkabau- West Sumatra, and other local-based folk stories tended to be scored higher in their own ethnic group location, despite the fact that some of these kinds of stories were also interesting to students of other ethnic groups.

It was also interesting that in a location where the use of Bahasa Indonesia was dominant, children tended to have better comprehension of the content of a book, a higher percentage of completing all pages of the book, and less frequent encountering of words, idioms, and paragraphs that they did not understand. Some locally-oriented words and idioms used by writers in their books were not fully understood by students from other locations, e.g.. Javanese idioms by Sundanese children, Balinese words by Malay children of Medan and Jambi, and the like.

These findings support the hypothesis that the degree of "interestingness" and readability of a book are to some extent relative to who are the readers, from which cultural background they come, and how much culturally-specific content appears in the book. It is also clear that the closer the content of a book is with the life of children, the more interesting the book is to those children.

Diversity of student multicultural backgrounds is also accommodated by use of reading books with local content, which each province has the authority to choose.

Like all other schoolbooks to be used at schools, local content reading books should also be legalized by the Director General of PSE. However, the evaluation process for these books is done by each Kanwil (Provincial Office of MOEC). Data at the Directorate of Educational Facilities shows that most of local content reading books contain local stories, traditions, customs, heroes, and languages. Unfortunately, not all provinces could provide these kinds of books because of the lack of quality writers and the absence of publishers.

As mentioned earlier, the textbook project for junior secondary schools is managed under several innovative policies, one of which is that textbook procurement is carried out by Kanwils in 27 provinces. From the list of textbook titles that have been approved by the NTEC, each Kanwil has to nominate at least three titles for each subject and grade. Of these titles, each Kanwil has eventually to come up with only one title to be produced for distribution to all junior secondary schools in its province. Preceding the process of book selection, each Kanwil has to conduct a survey of a limited sample of students and teachers in at least 10 schools to identify which book is the most appropriate for purchase by the project. The criteria for selection are no longer related to quality of content, language, graphical presentation, and security aspects; rather the main criterion is local relevance to a particular province.

From the perspective of book writers, this would imply that from the very beginning they should consider the targeted provinces in which their books are expected to be bought by the project. This policy is consistent with research findings in the area of multicultural education in that the ability of teachers and students to use and to benefit from the textbooks depend on the extent to which the textbooks meet the needs of users and are in line with their characteristics (Gollnick and Chinn, 1991).

It is clear that schoolbook provision (both textbooks and reading books) in Indonesia is moving toward incorporating multicultural aspects of student life and their diverse cultural environments. This development is still in its early stage, and the pendulum is gradually moving from the extremely "national- oriented" education (including educational books) which had long neglected the reality of diverse cultures--because of political reasons which were judged as valid at the time--toward education that recognizes local cultures, traditions, styles, and levels of community development.

Notes

1. SD (Sekolah Dasar) is general primary school under the Ministry of Education and Culture and Ministry of Home Affairs; MI (Madrasah Ibtidaiyah) is the Islamic Primary School under the Ministry of Religious Affairs.
2. According to this study, at primary school the correlations between pupils' reading behavior and the frequency of borrowing books from school library, number of books in school library, and presence of school library were .18, .07, and .05, respectively. For junior secondary school, the correlations between student reading behavior and the frequency of visiting school library and number of books in school library were .12 and .01.
3. The Pancasila (The Five Basic Principles) is the official state ideology which consists of five principles which are formulated in the preamble of the UUD 1945 (The National Constitution) as follows: (1) Belief in the One and only One God; (2) Just and civilized humanity; (3) The unity of Indonesia; (4) Democracy guided by the inner wisdom of deliberations and representatives; and (5) Social justice for all the Indonesian people.
4. In this project, I was assigned as a Coordinator of monitoring and evaluation, covering 27 provinces.

References

Cousin, Patricia Tefft. 1989. *Content Area Textbooks: Friends or Foes?* ERIC Digest. ED321249.

Direktorat Sarana Pendidikan, 1995. *Rencana Proyek Pengembangan Buku dan Minat Baca dalam Rangka Wajib Belajar 9 Tahun.* Jakarta: Ditsardik, Ditjen Dikdasmen.

Gollnick, Donna M., and Chinn, Philip C. 1991. *Multicultural Educational for Exceptional Children*. ERIC Digest # E498, ED333620, May 1991.

Heyneman, S.P., Fareel, J., & Sepulveda-Stuarto, M. 1981. "Textbooks and Achievement: What We Know?" *Journal of Curriculum Studies*, 13:3.

Hoge, John D. 1986. *Improving the Use of Elementary Social Studies Textbooks*. ERIC Digest, ED274582, September 1986.

Lockeed, M. & Verspoor, A. 1990. *Improving Primary Education in Developing Countries: A Review of Policy Options*. Washington, D.C.: World Bank.

Patrick, John J. 1988. *High School Government Textbooks*. ERIC Digest, ED301532, December 1988.

Proyek Pengembangan Buku dan Minat Baca. 1996. *Buku Panduan Pemasyarakatan Buku dan Minat Baca*. Jakarta: Ditjen Dikdasmen.

Supriadi, Dedi. 1996a. Satu Buku Satu Siswa. *Suara Karya*, Mei 1996

Supriadi, Dedi. 1996b. *Laporan Hasil Pemilihan Buku Bacaan SD/MI 1996*. Direktorat Sarana Pendidikan, Ditjen Dikdasmen, Depdikbud.

Supriadi, Dedi. 1996c. *Promosi Buku dan Minat Baca di Sekolah*. Jakarta: PBMB, Ditjen Dikdasmen, Depdikbud.

Supriadi, Dedi. 1995. *Laporan Hasil Pemilihan Buku Bacaan SD/MI 1995*. Direktorat Sarana Pendidikan, Ditjen Dikdasmen, Depdikbud.

World Bank. 1996. *Indonesia: Book and Reading Development Project, Loan 3887-IND. Draft Aide-Memoire*. May 1996.

World Bank. 1995a. *Indonesia: Book and Reading Development Project, Staff Appraisal Report*. May1995.

World Bank. 1995b. *Indonesia: Book and Reading Development Project, Loan 3887-IND. Draft Aide-Memoire*. September 1995.

World Bank. 1994. *Indonesia: Stability, Growth, and Equity in Repelita VI*. Country Department III, East Asia and Pacific Region. Washington D.C.: World Bank.

World Bank. 1992. *Indonesia: Book and Reading Development Study*. Washington, D.C.: World Bank.

World Bank. 1989. *Indonesia: Basic Education Study*. Washington D.C.: World Bank.

About the Author

Dedi Supriadi

Faculty Member
Graduate School of Education
Institute of Teaching and Educational Sciences (IKIP)
Bandung, West Java, INDONESIA

Phone: 062-22-707486
Fax: 062-22-707486 & 062-22-2001197
E-mail: supriadi@indo.net.id

Dedi Supriadi received his Ph.D. (Cum Laude) in Educational Psychology and Counseling from IKIP Bandung, Indonesia. He is seconded to the office of Minister of Education and Culture of the Republic of Indonesia as a speech writer and program developer, assistant to the Director Generale of Primary and Secondary of Education (DGSPE), Chairman of Textbook Evaluation Committee at the DGSPE, Chairman for Research and Development Division of the Association of Reading Interest Promotion (Perhimpunan Masyarakat Gemar Membaca, PMGM), and currently is Deputy Director for Academic Affairs at the Graduate School of Education, IKIP Bandung, Indonesia and editor of Mimbar Pendidikan, an Educational Journal published by the IKIP. Dr. Dedi Supriadi has written several books on counseling, creativity, book writing and evaluation, higher education in Indonesia, and teacher education.

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin
John Covaleskie
Northern Michigan University
Alan Davis
University of Colorado, Denver
Mark E. Fetler
California Commission on Teacher Credentialing
Thomas F. Green
Syracuse University
Arlen Gullickson
Western Michigan University
Aimee Howley
Ohio University
William Hunter
University of Calgary
Daniel Kallos
Umeå University
Thomas Mauhs-Pugh
Green Mountain College
William McInerney
Purdue University
Les McLean
University of Toronto
Anne L. Pemberton
apembert@pen.k12.va.us
Richard C. Richardson
Arizona State University
Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning
Michael Scriven
scriven@aol.com
Robert Stonehill
U.S. Department of Education
David D. Williams

Greg Camilli
Rutgers University
Andrew Coulson
a_coulson@msn.com
Sherman Dorn
University of South Florida
Richard Garlikov
hmwkhelp@scott.net
Alison I. Griffith
York University
Ernest R. House
University of Colorado
Craig B. Howley
Appalachia Educational Laboratory
Richard M. Jaeger
University of North Carolina-- Greensboro
Benjamin Levin
University of Manitoba
Dewayne Matthews
Western Interstate Commission for Higher Education
Mary McKeown-Moak
MGT of America (Austin, TX)
Susan Bobbitt Nolen
University of Washington
Hugh G. Petrie
SUNY Buffalo
Anthony G. Rud Jr.
Purdue University
Jay D. Scribner
University of Texas at Austin
Robert E. Stake
University of Illinois--UC
Robert T. Stout
Arizona State University

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Maria Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 7700 times since March 23, 1999

Education Policy Analysis Archives

Volume 7 Number 8 March 23, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

BEST COPY AVAILABLE

Scholastic Achievement and Demographic Characteristics of Home School Students in 1998

Lawrence M. Rudner
ERIC Clearinghouse on Assessment and Evaluation
College of Library and Information Services
University of Maryland, College Park

Abstract

This report presents the results of the largest survey and testing program for students in home schools to date. In Spring 1998, 20,760 K-12 home school students in 11,930 families were administered either the Iowa Tests of Basic Skills (ITBS) or the Tests of Achievement and Proficiency (TAP), depending on their current grade. The parents responded to a questionnaire requesting background and demographic information. Major findings include: the achievement test scores of this group of home school students are exceptionally high--the median scores were typically in the 70th to 80th percentile; 25% of home school students are enrolled one or more grades above their age-level public and private school peers; this group of home school parents has more formal education than parents in the general population; the median income for home school families is significantly higher than that of all families with children in the United States; and almost all home school students are in married couple families. Because this was not a controlled experiment, the study does not demonstrate that home schooling is superior to public or private schools and the results must be interpreted with caution. The report clearly suggests, however, that home school students do quite well in that educational environment.

By current estimates, there are between 700,000 and 1,200,000 students enrolled in home schools in the United States. Further, by all accounts, the movement has been growing steadily over the past few years (Lines, 1998). Yet, there is very little scientific literature concerning the population of home school students or even large samples of home school students.

This study describes the academic achievement levels and some basic demographic characteristics of a large sample of students and their families. While the academic levels of home school students are described in terms of public and private school norms, this study is not a comparison of home schools with public or private schools. Such comparisons would be fraught with problems. Home schooling is typically one-on-one. Public schools typically have classes with 25 to 30 students and an extremely wide range of abilities and backgrounds. Home school parents are, by definition, heavily involved in their children's education; the same, unfortunately, is not true of all public or private school parents. Home schools can easily pace and adapt their curriculum; public and private schools typically have a mandated scope and sequence. The list of differences could continue.

This study seeks to answer a much more modest set of questions: Does home schooling tend to work for those who chose to make such a commitment? That is, are the achievement levels of home school students comparable to those of public school students? Who is engaged in home schooling? That is, how does the home school population differ from the general United States population?

Methods

Bob Jones University Press Testing and Evaluation Service provides assessment services to home school students and private schools on a fee-for-service basis. In Spring 1998, 39,607 home school students were contracted to take the Iowa Tests of Basic Skills (ITBS; grades K-8) or the Tests of Achievement and Proficiency (TAP; grades 9-12). Students were given an achievement test and their parents were asked to complete a questionnaire entitled "Voluntary Home School Demographic Survey." A total of 20,760 students in 11,930 families provided useable questionnaires with corresponding achievement tests. The achievement test and questionnaire results were combined to form the dataset used in this analysis.

This section provides descriptions of the achievement measures, the questionnaires, the Bob Jones University Press Testing and Evaluation Service, and the procedures used to develop the dataset.

Iowa Tests of Basic Skills (ITBS)

Home schooled students in Grades K-8 took the Iowa Tests of Basic Skills (ITBS) Form L, published by Riverside Publishing Company, a subsidiary of Houghton Mifflin. Developed by University of Iowa professors, the tests were designed and developed to measure skills and standards important to growth across the curriculum in the nation's public and private schools.

The ITBS reflects more than 50 years of test development experience and research on measuring achievement and critical thinking skills in Reading, Language Arts, Mathematics, Social Studies, Science, and Information Sources. The scope and sequence of the content measured by the ITBS were developed after careful review of national and state curricula and standards, current textbook series and instructional materials, and research (Riverside, 1993).

All items were tried out and tested for ethnic, cultural, and gender bias and fairness prior to the development of the final form of the tests. Data on a nationally representative sample of public and private schools were collected in 1992 and used to form the initial national norms. The norms were updated in 1995 by Riverside. This study used these 1995 spring norms.

Tests of Achievement and Proficiency (TAP)

Home schooled students in Grades 9-12 took the Tests of Achievement and Proficiency (TAP), Form L, also published by Riverside Publishing Company. The TAP was designed and developed to measure skills and standards important to growth across the high school curriculum. Like the ITBS, the TAP scope and sequence were developed after careful review of national and state curricula and standards, and current textbook series and instructional materials. Developed as an upward extension of the ITBS, the specifications, format, and design of the TAP tests are similar to that of the ITBS. TAP is fully articulated with the Iowa Tests of Basic Skills (ITBS) Form L (Riverside, 1993).

Background Questionnaires

Background questionnaires were designed by the staff of the Home School Legal Defense Association (HSLDA). Questions were determined by reviewing the questions in previous surveys, prioritizing them, and selecting only those that were most germane to the objectives of the study. Where possible, questions and responses were made to match those used by the U.S. Census, U.S. Department of Labor and the National Assessment of Educational Progress to facilitate comparisons of home school students with students nationwide.

HSLDA designed the survey to be much shorter than previous survey instruments. They also sought to pose all questions in an objective format, rather than a constructed response format. In keeping with this approach, HSLDA worked with National Computer Systems to design forms to be computer scanable, thereby removing the need for manual data processing.

Bob Jones University Press Testing and Evaluation Service

The Bob Jones University (BJU) Press Testing and Evaluation

Service is the largest and oldest of four organizations providing home school families access to standardized achievement tests. The Testing Service began offering the Iowa Tests of Basic Skills and Tests of Achievement and Proficiency in 1984. In subsequent years they added other helpful tools including practice materials, a personality inventory, and diagnostic tests. In 1993, the Stanford Achievement Test series was added as BJU Press assumed the testing that the Home School Legal Defense Association had been providing for its members. Since that time, a full range of writing evaluations (grades 3-12) and a career assessment have been added to the growing number of evaluation tools offered by the Testing Service.

Just as home school families were the impetus behind the start of the Testing Service, home school families continue to be the largest sector utilizing the service. However, there are also a number of private schools that have chosen to use the services provided. Testing is provided for students throughout the United States and Canada, as well as many foreign countries.

The BJU Press Testing and Evaluation Service sends testing materials to qualified testers who administer the tests and return them to the Testing Service for scoring. The results are then returned to the parent. Many parents test primarily for their own information to verify that their home schooled students are progressing academically at a normal pace. Other parents use the results to meet a state testing requirement or to provide documentation when they choose to return their students to a public or private school setting.

Data Generation Procedures

The following steps were followed to produce the data set:

1. Parents contracted with Bob Jones University to be administered the Iowa Tests of Basic Skills or the Tests of Achievement Proficiency (39,607 students in probably 22,000 families).
2. Bob Jones certified test administrators, many of whom were the students' parents.
3. BJU sent questionnaires and answer forms to the test administrators.
4. Tests and questionnaires were returned to BJU. BJU bundled the tests and sent them to Riverside Publishers for machine scoring. BJU bundled the questionnaires and sent them to National Computer Systems for scanning. Unlike in previous studies, the parents did not know their scores ahead of time.
5. Electronic copy of the 23,415 test results and 23,311 questionnaire results were sent to the author of this report. These sets were merged to provide 20,900 cases with matching identification numbers. In order to weight by state public school enrollment, 140 cases with missing state data were dropped. A total of 20,760 students formed the initial dataset used in the study. After we formed the dataset with 20,760 students, we asked for the remainder of the 39,607 achievement test scores. We were informed that it would not be possible to disaggregate

the remaining home school students from students in private schools also contracting testing services.

Characteristics of Home School Students and Families

This section provides a description of home school students and their families based on the 20,790 respondents to our questionnaire. The distribution of students by state, gender, age, race, parent marital status, family size, mother's religion, parent education, family income, television viewing, money spent on educational materials, and other demographic characteristics are identified and, where possible, compared to national figures.

State

As shown in Table 2.1, respondents came from each of the fifty states. Several states, including Ohio, Georgia, and Virginia, have exceptionally high representation given their size. This is probably due to the fact that these states require testing of home school students. To reduce the effects of these and other overrepresented states, the data were weighted in all subsequent analyses by the number of public school students in each state. While we would have preferred to weight by the number of home schooled students in each state, such data are not available for all 50 states (Lines, 1998).

Table 2.1
Participating Home
School Students Classified by State

State	Freq.	Percent of sample	State	Freq.	Percent of sample
AK	61	.3%	MT	112	.5
AL	181	.9	NC	972	4.7
AR	42	.2	ND	100	.5
AZ	201	1.0	NE	126	.6
CA	815	3.9	NH	176	.8
CO	810	3.9	NJ	324	1.6
CT	54	.3	NM	189	.9
DC	17	.1	NV	53	.3
DE	28	.1	NY	942	4.5
FL	860	4.1	OH	2484	11.9
GA	1547	7.4	OK	382	1.8
GU	10	.0	OR	67	.3
HI	112	.5	PA	532	2.6
IA	234	1.1	PR	8	.0
ID	28	.1	RI	32	.2
IL	451	2.2	SC	579	2.8
IN	533	2.6	SD	27	.1
KS	319	1.5	TN	322	1.5
KY	163	.8	TX	1126	5.4
LA	551	2.7	UT	35	.2
MA	343	1.6	VA	1608	7.7
MD	196	.9	VI	2	.0
ME	109	.5	VT	59	.3
MI	523	2.5	WA	787	3.8
MN	794	3.8	WI	246	1.2
MO	361	1.7	WV	92	.4
MS	25	.1	WY	40	.2

Student Age and Gender

Table 2.2 shows the distribution of the respondents by gender and age. About 50.4% or 10,471 of the respondents were females; 49.6% (10,319) were males. These figures are comparable to that of the population of 3 to 34 years old enrolled in school (see U.S. Bureau of the Census, 1998, Table A-2). Some 51.4% of school enrollees nationally are male. The percentages are comparable at all age levels.

Table 2.2
Participating Home School Students
Classified by Gender and Age

	Age at time of testing (in years)											
	6	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
Females	507	879	1148	1318	1301	1248	1049	936	774	516	264	119
	56.1%	51.7%	50.2%	49.2%	52.4%	50.6%	47.2%	50.5%	50.7%	51.0%	49.3%	57.5%
Males	397	820	1141	1360	1181	1216	1174	918	754	495	271	88
	43.9%	48.3%	49.8%	50.8%	47.6%	49.4%	52.8%	49.5%	49.3%	49.0%	50.7%	42.5%
Total	904	1699	2289	2678	2482	2464	2223	1854	1528	1011	535	207

Student Grade

Home school student grade placement was identified by their parents, presumably based on the grade level of the instructional materials. That grade was used by BJU to determine the test levels and used in this report as a grouping variable. Tables 2.3 shows the distribution of respondents and the nation by grade. There is a large difference in the proportions of high school (grades 9-12) home school students and the nation. Compared to the national data, a relatively small percentage of home school students are enrolled in high school. Possible reasons for this lower participation for high school students may be the relative newness of the home school movement, early graduation from high school, and possibly a desire on the part of some home school parents to enroll their children in a traditional high school. The distributional differences for students in grades 1 through 8 are minor.

Table 2.3
Home School Students Classified by Grade with Percents and National School Percents

	Grade											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Home school	7.4% (1504)	10.6% (2153)	14.1% (2876)	12.9% (2625)	12.6% (2564)	11.9% (2420)	10.3% (2087)	8.8% (1801)	5.7% (1164)	3.8% (775)	1.6% (317)	0.3% (66)
Nation	9.1%	8.8%	8.9%	8.7%	8.6%	8.7%	8.7%	8.4%	9.0%	7.9%	7.1%	6.3%

National data: US Census, 1997b, Table 254.

Student Race

Table 2.4 shows the racial distribution of home school students in 1998 and for the students enrolled in elementary and secondary public and private schools nationally in 1994. The distributions are quite different. The vast majority of home schooled children are non-Hispanic White. The largest minority groups for home school students (not shown in the table) are American Indians and Asian students who comprise some 2.4% and 1.2% of the home school students, respectively.

Table 2.4
Racial Distribution of Home School Students And the Nation, in Percents

	White (not Hispanic)	Black (not Hispanic)	Hispanic	Other
Home school	94.0%	0.8%	0.2%	5.0%
Nationwide	67.2%	16.0%	13.0%	3.8%

(National data: USDE, 1996; Indicator 27)

Marital Status

The great majority of home school students are in married couple families. In contrast, only 72% of the families with at least one child enrolled in school nationwide are in married couple families (Bruno and Curry, 1997, Table 19).

Table 2.5
Home School Students Classified by Parents' Marital Status

Marital Status	Frequency	Percent
Divorced	80	0.7%
Single (never married)	44	0.4
Married	11,335	97.2
Separated	131	1.1
Widowed	55	0.5
Missing data	16	0.1
	11,661	100.0%

Children at Home

Table 2.6 shows the distribution of children in home school families and families with children under 18 nationwide. On average, home school students are in larger families. Nationwide, most families with school-age children (79.6%) have only 1 or 2 children with a mean of about 1.9 children per family. Most home school families (62.1%) have 3 or more children with a mean of about 3.1 children per family.

Table 2.6
Home School Families Classified

by Family Size with National Comparison

Home School Families		Nationwide	
Number of Children	**Percent**	**Number of Children**	**Percent**
1	8.3%	1	40.8%
2	29.6	2	38.8
3	28.6	3	14.3
4	18.6	4 or more	6.1
5	8.4		
6	3.9		
7 or more	2.6		

National Data: US Census, 1997a, Table 77

Mother's Religion

We asked the home school families to identify the religious preference of each student's mother by selecting from a list of 27 religions. As shown in Table 2.7, the largest percentage of mothers identified themselves as Independent Fundamental, Baptist, Independent Charismatic, Roman Catholic, Assembly of God, or Presbyterian. The religious preference of the father was the same as that of the mother 93.1% of the time.

Table 2.7
Home School Students Classified by Mother's Religion

	Frequency	Percent
Independent Fundamental	5,119	25.1%
Baptist	5,072	24.4
Independent Charismatic	1,681	8.2
Roman Catholic	1,106	5.4
Assembly of God	838	4.1
Presbyterian	772	3.8
Reformed	685	3.4
Other Protestant	500	2.5
Pentecostal	459	2.2
Methodist	420	2.1
Lutheran	353	1.7
Other Christian	2,213	10.9
Other	1,572	6.2
Total	20,790	100.0%

Parent Academic Attainment

As shown in Table 2.8, home school parents have more formal education than the general population. While slightly less than half of the general population attended or graduated from college, almost 88% of home school students have parents who continued their education after high school.

Table 2.8
Distribution of Home School Students and Students Nationally Classified by Parent Academic Attainment

	Percent						
	Did not finish high school	**High school graduate**	**Some college, no degree**	**Associat degree**	**Bachelo degree**	**Masters degree**	**Doctorate**
Home school fathers	1.2%	9.3%	16.4%	6.9%	37.6%	19.8%	8.8%
Nation males	18.1	32.0	19.5	6.4	15.6	5.4	3.1
Home school mothers	0.5	11.3	21.8	9.7	47.2	8.8	0.7
Nation females	17.2	34.2	20.2	7.7	14.8	4.5	1.3

National data: U.S. Census (1996; Table 8)

Family Income

National data on family income are available for 1995. As shown in Table 2.9, home school families span all income levels. On average, home school families have a higher income level than do families with children nationwide and all families nationwide. The median family income level for home school families in 1997 is about $52,000. The median income for families with children in 1995, nationwide, was about $36,000.

Table 2.9
Distribution of Family Income for Home School Families, Families with Children Nationwide, and All Families Nationwide by Income Levels, in Percents.

	Home school	Families with children	All families
Less than $10,000	0.8%	12.6%	10.5%
$10,000 to $14,999	1.5	8.0	8.5
$15,000 to $19,999	2.2	6.1	6.8
$20,000 to $24,999	3.9	7.6	8.4
$25,000 to $29,999	4.9	7.5	7.8
$30,000 to $34,999	8.5	7.5	7.6
$35,000 to $39,999	8.1	7.1	7.0
$40,000 to $49,999	16.0	11.3	11.0
$50,000 to $74,999	32.5	18.4	18.1
$75,000 and over	21.6	13.8	14.3

National data: Bruno and Curry (1997, Table 19)

Television Viewing

The National Assessment of Educational Progress collects information on the television viewing habits of fourth-graders. Home school fourth-graders and fourth-graders nationally differ markedly in terms of television viewing. Home school students rarely watch more than 3 hours of television per day; nearly 40% of the students nationwide watch that much television.

Table 2.10
Fourth-grade students Classified by Hours of Television Viewing

Percent of students

	6 or more hours per day	4 to 5 hours per day	2 to 3 hours per day	1 hour or less per day
Home school	0.1%	1.6	33.1	65.3
Nationwide	19.0%	19.5	36.4	25.1

National data: NAEP Math 1997

Computer Use

The *Condition of Education* provides a tabulation of the percent of students nationwide who report using a computer by frequency of use for 4th, 8th, and 11th graders in 1996. At each grade level, the distribution of computer use in 1998 by home school students is different from that of the nation in 1996. At each of these three grade levels, much larger percentages of home school students never use a computer. At the fourth-grade level, a much larger percent of home school students use a computer every day.

Table 2.11
Computer Use among Home School Students and Students Nationwide in Grades 4, 8, and 11, in Percent

	Grade 4		Grade 8		Grade 11	
	Home school	Nationwide	Home school	Nationwide	Home school	Nationwide
Never	28.2%	11.4%	37.1%	23.3%	40.5 %	16.0%
Less than once a week	29.4	16.3	28.9	29.2	28.9	34.2
Several times a week	21.6	62.5	18.0	30.7	17.5	31.8
Every day	20.8	9.9	16.0	16.7	13.1	18.1

National Data: Snyder and Wirt, 1998, Indicator 3

Money Spent on Educational Materials

The amount of money spent in 1997 on home school education for textbooks, lesson materials, tutoring and enrichment services, and testing ranged from less than $200 to more than $2000. As shown in Table 2.12, the median amount of money spent was about $400.

Table 2.12
Home School Students Classified by Money Spent On Home School Education in 1997

Amount	Frequency	Percent
<$200	3,718	17.9%
200-399	7,035	33.8
400-599	4,467	21.5
600-799	1,962	9.4
800-999	985	4.7
1,000-1,599	1,630	7.8
1,600-1,999	247	1.2
>2,000	411	2.0
Missing	336	1.6
Total	20,790	100.0%

Other Demographic Characteristics

Compared to the nation, a much larger percentage of home school mothers are stay-at-home mothers not participating in the labor force. Some 76.9% of home school mothers do not work for pay. About 86.3% that do work do so part time. Nationwide, in 1996, only 30% of married women with children under 18 did not participate in the labor force (US Dept of Census, 1997a, Table 632).

A very large percentage of home school parents are certified to teach. Some 19.7% of the home school mothers are certified teachers; 7.1% of fathers. Almost one out of every four home school students (23.6%) has at least one parent who is a certified teacher.

Only 7.7% of the respondents were enrolled in a full-service curriculum program, i.e., a program that serves students and their parents as a "one-stop" primary source for textbooks, materials, lesson plans, tests, counseling, evaluations, record keeping, and the like for the year's core required subjects such as language, social studies, mathematics, and science.

Academic Achievement

The complete batteries of The Iowa Tests of Basic Skills (ITBS) and the Tests of Achievement and Proficiency (TAP) were used to assess student achievement in basic skills. The ITBS was used for home school students in Grades K-8; the TAP for students in grades 9-12. Almost all students took Form L; a handful took parallel Form K.

Achievement test batteries like the ITBS and TAP are a collection of tests in several subject areas that have been standardized and normed. Norms for all tests within these test batteries are based on the same group of students at each grade level. Such norms allow students to be compared with other students and groups to be compared with other groups.

The primary purpose of the ITBS and TAP is to assess the academic achievement of students in public and private schools. Consequently, much of the test development effort is devoted to identifying the content to be covered by these batteries. Riverside Publishers follow a four step process: 1) content specifications, 2) editorial review, 3) pilot testing, and 4) national norms development and updating.

The first and most critical step is developing content specifications and writing test items. This step involves the experience, research, and expertise of a large number of professionals representing a wide variety of specialties in the education community. Specifications are developed which outline the grade placement and emphasis of skills. These specifications draw heavily on an analysis of textbooks, research studies, nationally developed subject matter standards, and national curriculum committees.

Once the items have been developed and pilot tested, the final forms of the tests are developed and administered to large standardization samples to gather normative data and to develop scales.

The spring standardization sample for the 10 levels of the ITBS consisted of approximately 137,000 students from public schools, Catholic schools and private non-Catholic schools. The public school sample was stratified to assure adequate representation based on geographic region, district enrollment, socioeconomic status of the district. The Catholic school sample was stratified on geographic region and diocese enrollment. The non-Catholic private school sample was stratified on region and school type. The spring standardization sample for the four levels of the TAP consisted of approximately 20,000 students stratified on the same variables. National norms were developed based on the combined weighted distributions of all three school types: public, Catholic and non-Catholic private. Catholic/private school norms were developed based on the combined weighted distributions of the latter two groups. For simplicity, the combined public, Catholic and non-Catholic private school norms are referenced in this report as national norms or public/private school norms.

The data from the standardization sample are used to develop a variety of reporting scales, such as percentiles and grade equivalent scores. The analyses in this report rely primarily on the Developmental Standard Score (DSS) scale developed by Riverside Publishers. The DSS is a number that describes a student's location on an achievement continuum that spans grades K through 12. Table 3.1 shows the median DSS and median age that corresponds to each grade level in the national standardization sample. The DSS scale shows that the average annual growth in DSS units decreases each year.

Table 3.1
Median Developmental Scaled Scores and Median Age for the ITBS/TAP Spring National Standardization Sample

Grade	K	1	2	3	4	5	6	7	8	9	10	11	12
DSS	130	150	168	185	200	214	227	239	250	260	268	275	280
Age	6.1	7.2	8.2	9.3	10.2	11.2	12.2	13.2	14.2	15.2	16.2	17.2	18.1

Source for age medians: Drahozal (1998, personal communication)

This same DSS scale is used for all tests and levels of the ITBS and TAP. The main advantages of the DSS are that it mirrors reality well, spans all grade levels, and provides a quasi equal interval scale which has a variety of attractive statistical properties. Most importantly, DSS scores can be compared to each other and can be meaningfully averaged.

The main disadvantage of DSS scores is that they have no built-in meaning. Reference points are needed to interpret DSS scores. "Grade level" is one possible reference point. A DSS score of 170 in reading, for example, is about equal to the typical reading score for second-grade students in public and private schools in the spring of the year. A more refined reference is the percentile score that corresponds to each DSS score. The 170 in *reading*, for example, corresponds to the 54th percentile of second graders. That is, this score is better than the score received by 54 percent of the second graders using the 1995 spring norms.

The reader should note that while all tests of the ITBS/TAP have the same median DSS score at each grade level, the distributions within each subject area vary. A DSS score of 310 for a tenth grader in *reading*, for example, corresponds to the 87th percentile. A DSS score of 170 in *mathematics* for a tenth grader would place the student at the 79th percentile.

Percentiles are always defined in terms of a grade level. This can be problematic when analyzing data for home school students. In this study, 24.5% of the home school students were one or more grades above the grade usually associated with that student's age (see Table 3.2). A strong case can be made that rather than using the percentile corresponding to the enrolled grade, as we did in this study, one should use the percentile associated with the student's nominal grade, i.e., the grade usually associated with the student's age. The argument is that a 10-year-old home school student enrolled in 5th grade should be compared to his age peers in 4th grade. The counter argument is that the percentiles already consider the fact that students are not always in their nominal grade since the standardization sample had students above and below grade level. We initially analyzed the data both ways. Rather than expose our analysis to criticism, we chose to take the more conservative route by employing the enrolled grade.

While very meaningful, percentiles do not provide a complete picture of a student's or group's academic performance. In this study, we used grade equivalent scores as an additional reference point for interpreting DSS scores. A grade equivalent score approximates a child's development in terms of grade and month within grade. A DSS reading score of 170 can be viewed as the typical DSS score earned by students in the ninth month of the second grade or a GES score of 2.9. Just as the percentile associated with a DSS scores varies by subtest, so do the properties of GES scores vary across subjects.

Grade Equivalent Scores are particularly useful for estimating a student's developmental status in terms of grade. But, these scores must be interpreted carefully. An GES Score of 6.3 in reading for an 9 year old in the 3rd grade, for example, clearly indicates that the third grader is doing well. This does not, however, mean that the third grader belongs in the 6th grade. It only means that the third grader can read as well as a sixth grader.

The usual interpretation of a Grade Equivalent Score of 6.3 for a third grader is that this third grade student can read third grade material as well as a sixth grader can read third grade material, not that he or she can read sixth grade material. The DSS of the ITBS/TAP, however, is unique. The DSS scales were developed by administering the same special scaling test to students in grades K-3, another common scaling test to students in grades 3 to 9, and another to students in grades 8-12. Thus, in the scaling study, the third graders did take the same test as the sixth graders in each subject area.

Grade Placement

Home school students are able to progress through instructional material at the student's rate. Thus, it is easy for home school students to be enrolled one or more grades above their public and private school-age peers. To evaluate the frequency of advanced placement, we compared students' enrolled and nominal grades. The enrolled grade was identified by the parents and used to determine the ITBS/TAP level. The nominal grade is the public school grade in which the student would normally be enrolled in based on the child's month and year of birth.

As shown in Table 3.2, almost one fourth of the home school students (24.5%) are enrolled one or more grades above their nominal grade. While comparable figures nationally do not exist, one research director in a large school district estimated that less than 5% of their students are enrolled above grade level.

Table 3.2
Home School Students Classified

by Discrepancy Between
Enrolled and Nominal Grade

Enrolled minus Nominal Grade	Frequency	Percent
-2	58	0.3%
-1	1,019	5.1
0	13,931	69.8
+1	4,637	23.2
+2	199	1.0
+3	58	0.3

Percentages do not sum to 100% due to a small percentage of students outside this range.

Overall Achievement

Table 3.3 shows the median scaled score (DSS score) for home school students on the *Composite with Computation, Reading Total, Language, Mathematics Total with Computation, Social Studies*, and *Science* subtest scores by grade. The corresponding percentiles shown in the table are the within grade percentile scores for the nation that correspond to the given scaled scores. For example, home school students in Grade 3 have a median composite scaled score of 207 which corresponds to the 81st percentile nationwide. The median home school student in third grade out- performs 81% of the third graders nationwide. As an additional comparison, we provide the national median for each grade in the last column. By definition this is the 50th percentile of students nationwide.

Table 3.3
Median Scaled Scores (corresponding national percentile) by Subtest and Grade for Home School Students

Grade	N	Composite	Reading	Language	Math	Soc. Stud.	Science	National Median
1	1504	170 (91)	174 (88)	166 (82)	164 (81)	166 (80)	164 (78)	150 (50)
2	2153	192 (90)	196 (89)	186 (80)	188 (85)	189 (81)	195 (86)	168 (50)
3	2876	207 (81)	210 (83)	195 (62)	204 (78)	205 (76)	214 (83)	185 (50)
4	2625	222 (76)	228 (83)	216 (67)	220 (76)	216 (68)	232 (81)	200 (50)
5	2564	243 (79)	244 (83)	237 (69)	238 (76)	236 (71)	260 (86)	214 (50)
6	2420	261 (81)	258 (82)	256 (73)	254 (76)	265 (81)	273 (84)	227 (50)
7	2087	276 (82)	277 (87)	276 (77)	272 (79)	276 (79)	282 (81)	239 (50)
8	1801	288 (81)	288 (86)	291 (79)	282 (76)	290 (79)	289 (78)	250 (50)
9	1164	292 (77)	294 (82)	297 (77)	281 (68)	297 (76)	292 (73)	260 (50)
10	775	310 (84)	314 (89)	318 (84)	294 (72)	318 (83)	310 (79)	268 (50)
11	317	310 (78)	312 (84)	322 (83)	296 (68)	318 (79)	314 (77)	275 (50)
12	66	326 (86)	328 (92)	332 (85)	300 (66)	334 (84)	331 (82)	280 (50)

It is readily apparent from Table 3.3 that the median scores for home school students are well above their public/private school counterparts in every subject and in every grade. The corresponding percentiles range from the 62nd to the 91st percentile; most percentiles are between the 75th and the 85th percentile. The lowest percentiles are in *Mathematics Total with Computation* subtest (labeled *Math* in the tables); the highest in *Reading Total*. While the grade-to-grade increase in national medians is 13 DSS points in the lower grades, the annual increase for home school students is about 16 points. These are exceptional scores and exceptional grade-to-grade gains.

As shown in Table 3.4, the same superiority of median scaled scores holds when comparing home school students to students enrolled in Catholic/Private schools. The Catholic/Private school percentiles corresponding to median scaled scores range from the 53rd percentile to the 89th percentile; most are between the 65th to 75th percentile. In every area and every grade, the median scores for home school students exceed the median scores of students enrolled in Catholic/Private schools.

Table 3.4

Median Scaled Scores of Home School Students (Corresponding Catholic/Private School Percentile) by Subtest and Grade

Grade	Composite	Reading	Language	Math	Soc. Stud.	Science
1	170 (89)	174 (86)	166 (80)	164 (80)	166 (73)	164 (75)
2	192 (88)	196 (84)	186 (74)	188 (81)	189 (81)	195 (85)
3	207 (74)	210 (74)	195 (55)	204 (71)	205 (69)	214 (80)
4	222 (72)	228 (72)	216 (58)	220 (69)	216 (56)	232 (76)
5	243 (71)	244 (72)	237 (60)	238 (68)	236 (60)	260 (82)
6	261 (71)	258 (71)	256 (58)	254 (65)	265 (72)	273 (77)
7	276 (72)	277 (77)	276 (63)	272 (70)	276 (68)	282 (73)
8	288 (72)	288 (75)	291 (65)	282 (68)	290 (68)	289 (67)
9	292 (63)	294 (70)	297 (61)	281 (56)	297 (63)	292 (59)
10	310 (71)	314 (81)	318 (71)	294 (57)	318 (72)	310 (66)
11	310 (63)	312 (72)	322 (69)	296 (56)	318 (67)	314 (63)
12	326 (74)	328 (81)	332 (71)	300 (53)	334 (74)	331 (72)

The relationship between median composite scaled scores for home school students, Catholic/Private school students, and the nation is shown in the Figure 1. At each grade level, the test performance of Catholic/Private school students is above the national performance levels, especially in the higher grade levels. Also at each grade level, the performance of home school students is above the performance levels of students enrolled in Catholic/Private schools. The differences between these groups are considerable. For example, the median score for 7th graders nationwide is 239; for Catholic/Private school students the median is 257; for home school students the median is 276. Another way to look at this chart is to examine the grades corresponding to a given composite score. A composite scale score of 250, for example, is typical of a home school student in Grade 6, a Catholic/Private school student in Grade 7 and students nationwide in the later stages of grade 8.

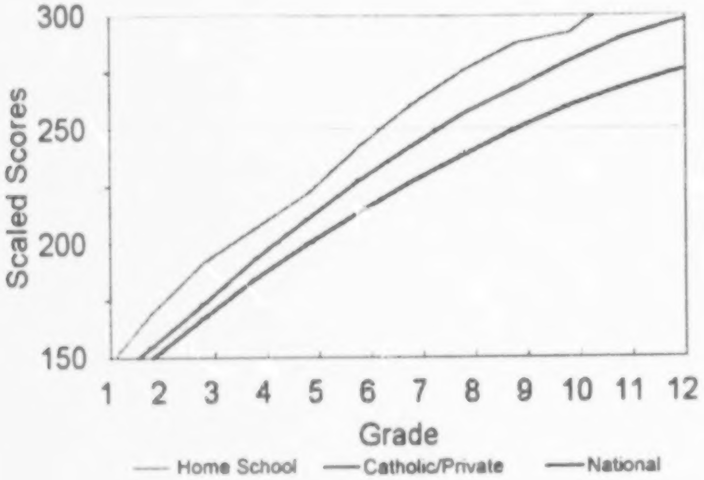


Figure 1. Academic Achievement of Home School, Catholic/Private and the Nation's Students

The Grade Equivalent Scores (GES) corresponding to the median DSS scaled scores for home school students are shown in Table 3.5. These GES scores indicate the performance levels of home school students in terms of student grade placement nationwide. The median composite scaled score for fourth-grade home school students, for example, is 217. This is comparable to the median score expected of students nationwide in the ninth month of fifth grade. Compared to students nationwide, the median fourth-grade home school student test performance is 1.1 grade equivalents above his public/private school peers. By 8th grade, the median performance of home school students on the ITBS/TAP is almost four grade equivalents above that of students nationwide. Similar trends hold for all subject areas.

The reader should recognize that the grade equivalent scale tends to magnify differences at the high school level and that the percentile scale is more meaningful in these higher grades. While 50% of eighth grade home school students have scores that are 4 grade equivalents above the public school median, so do some 20% of eighth grade students in public schools. The revealing statistics are the percentiles which are consistently high across grade levels and subject areas.

Table 3.5
Median Scaled Scores (corresponding Grade Equivalent Scores) by Subtest and Nominal Grade for Home School Students

Grade	Composite	Reading	Language	Math	Soc. Stud.	Science	National Median
1	170 (2.9)	174 (3.1)	166 (2.6)	164 (2.6)	166 (2.7)	164 (2.6)	150 (1.8)
2	192 (4.1)	196 (4.5)	186 (3.8)	188 (4.0)	189 (4.0)	195 (4.5)	168 (2.8)
3	207 (5.1)	210 (5.5)	195 (4.4)	204 (5.2)	205 (5.1)	214 (5.8)	185 (3.8)
4	222 (6.2)	228 (6.9)	216 (5.9)	220 (6.4)	216 (5.9)	232 (7.3)	200 (4.8)
5	243 (8.3)	244 (8.3)	237 (7.6)	238 (7.7)	236 (7.6)	260 (9.8)	214 (5.8)
6	261 (10.1)	258 (9.6)	256 (9.4)	254 (9.1)	265 (10.4)	273 (11.6)	227 (6.8)
7	276 (11.9)	277 (12.0)	276 (11.9)	272 (11.3)	276 (11.9)	282 (12.5)	239 (7.8)
8	288 (12.9)	288 (12.9)	291 (-)	282 (12.5)	290 (-)	289 (-)	250 (8.8)
9	292 (-)	294 (-)	297 (-)	281 (12.4)	297 (-)	292 (-)	260 (9.8)
10	310 (-)	314 (-)	318 (-)	294 (-)	318 (-)	310 (-)	268 (10.8)
11	310 (-)	312 (-)	322 (-)	296 (-)	318 (-)	314 (-)	275 (11.8)
12	326 (-)	328 (-)	332 (-)	300 (-)	334 (-)	331 (-)	280 (12.8)

(The - sign indicates the scaled scores are beyond the effective range for GES conversion.)

The grade equivalent score comparisons for home school students and the nation are shown in Figure 2. In grades one through four, the median ITBS/TAP composite scaled scores for home school students are a full grade above that of their public/private school peers. The gap starts to widen in grade five. By the time home school students reach grade 8, their median scores are almost 4 grade equivalents above their public/private school peers.

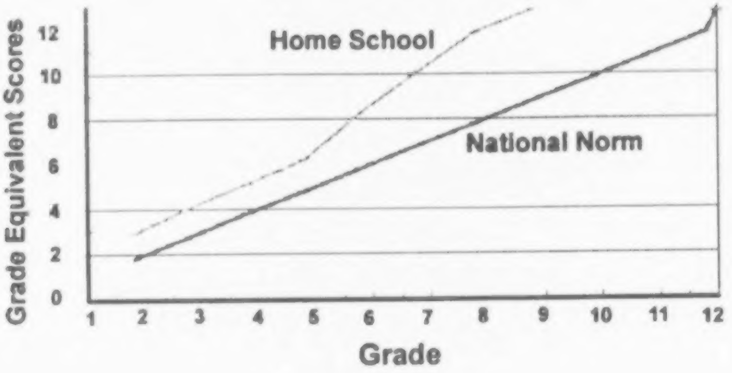


Figure 2. Home School Students Compared to the National Norm Group in Grade Equivalent Units

Years of Home Schooling

Almost half of the respondents (47%) indicated that they have been home schooled for each grade prior to their current grade, i.e., their entire academic life. Table 3.6 shows that students who are home schooled for their entire academic life do better than students who have been home schooled for only a few years *(F academic life =108.2; df=1,9750; p<.01)*. There is also a significant interaction between grade and years home schooled *(F=7.4; df=9,9750, p< .01)*, indicating that the effectiveness of home schooling varies with the student's grade. The differences are most meaningful starting in Grade 6.

[All F ratios reported here are from a two-way analysis of variance with composite scaled scores as the dependent measure, grade as a blocking variable, and one independent variable. Because the students are within families, the dataset was trimmed by randomly selecting one child from each family. Had the full dataset been used, the variance of the children within a family would have been artificially smaller than the variance of among children in the population of inference. This would have increased the risk of Type I error, showing significance when significance may not be so. To assure adequate cell sizes, the analyses were also restricted to Grades 1 through 10. A statistically significant difference only means that there is evidence of a difference in population values. The difference may be small and not meaningful. "n.s." is used to indicate not significant.]

One reviewer questioned whether this significant difference was due to life-long home schooling or was life-long home schooling serving as a proxy for parent education or income. The correlation of life-long home schooling and whether either parent has a college degree is .12, indicating there is some, but not a great deal of overlap between these variables. The correlation with income level was .02, indicating no relationship. Thus, whether a student is home schooled his or her entire life appears to be significantly related to achievement.

Table 3.6
Composite Scale Score Mean, Standard Deviation and Corresponding Percentile
by Number of Grades Home Schooled and Grade

Grade	1	2	3	4	5	6	7	8	9	10
Home schooled entire academic life										
Mean	170	195	208	224	244	265	278	291	300	314
sd	12	16	17	20	23	23	25	26	27	23
N	479	743	863	608	552	444	319	242	159	100
%ile	92	95	85	81	82	85	83	84	83	86
Home schooled some grades										
Mean	168	192	206	222	241	256	270	282	288	299
sd	11	15	18	20	24	26	27	30	30	32
N	221	428	616	666	681	688	628	608	436	287
%ile	90	92	82	79	79	78	77	78	73	75
Difference	2	3	2	2	3	9	8	9	12	15

[The percentiles (%ile) shown in this and the following tables are the within-grade percentiles corresponding to the mean composite scale scores, differences and ranges refer to differences in and ranges of mean composite scale scores, sd refers to standard deviation, N is the number of students within each cell.]

Enrolled in a Full-Service Curriculum

There is no significant difference in the mean composite scaled scores of home school students enrolled in a full-service curriculum and home school students not so enrolled. As shown in Table 3.7, the means are quite close at all grade levels *(F enrollment=.24; df=1,9750; n.s.)*.

Table 3.7

Composite Scale Score Mean, Standard Deviation and Corresponding Percentile by Full-service Curriculum Status and Grade

BEST COPY AVAILABLE

Grade	1	2	3	4	5	6	7	8	9	10
Not enrolled in a full-service curriculum										
Mean	170	194	207	223	243	260	272	284	291	302
sd	12	15	17	20	23	25	26	29	30	31
N	646	1109	1361	1214	1145	1042	847	771	495	320
%ile	92	94	83	80	81	81	79	79	76	78
Enrolled in a full-service curriculum										
Mean	167	199	209	220	241	256	272	286	289	306
sd	13	17	18	21	24	29	31	30	30	28
N	54	63	118	60	89	89	101	79	100	67
%ile	89	97	86	76	79	78	79	80	74	81
Difference	3	-5	-2	3	2	4	0	-2	-2	-4

Student Gender

There are no significant differences in the achievement levels of male versus female home school students *(F for gender=.01; df=1,9750; n.s.)*. As shown in Table 3.8, the means are virtually identical at all grade levels.

Table 3.8
Composite Scale Score Mean, Standard Deviation and Corresponding Percentile by Grade and Gender

Grade	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Males										
Mean	170	195	208	223	243	260	271	285	288	303
sd	12	15	18	19	23	25	26	30	33	33
N	355	576	749	639	600	597	479	428	294	181
%ile	92	95	85	80	81	81	78	80	73	78
Females										
Mean	169	193	207	223	242	260	274	284	293	303
sd	12	16	17	21	24	25	26	28	26	28
N	345	595	730	634	634	535	469	422	302	206
%ile	91	93	83	80	80	81	80	79	77	78
Difference	1	2	1	0	1	0	-3	1	-5	0

Money Spent on Educational Materials

There is a significant difference in the achievement levels of home school students depending on the amount of money spent per child on educational materials including textbooks, lesson materials, tutoring, enrichment services, and testing (see Table 3.9). At almost every grade level, students in families spending $600 or more outperform students in families spending less than $200 *(F for money spent=41.1; df=3,9585; p <.01)*. There is also a significant interaction between grade and money spent *(F=2.7; df=27,9585; p <.01)* indicating that the amount of money spent on education makes a bigger difference at the higher grade levels. The correlation between money spent on educational materials and income is significant (*r=.24, p <.01*), indicating that this effect may be due to family characteristics rather than expenditures.

Table 3.9
Composite Scale Score Mean, Standard Deviation and Corresponding Percentile by Money Spent on Educational Materials per Student and Grade

Grade	1	2	3	4	5	6	7	8	9	10
$600 or more										
mean	171	195	208	227	245	264	278	289	298	307
sd	11	16	17	21	23	25	25	30	27	32
N	152	236	408	329	317	306	289	260	226	147
%ile	93	95	85	84	83	84	83	83	81	81
$400-599										
mean	169	196	211	222	245	261	271	286	291	306
sd	13	15	17	19	22	25	26	25	31	30
N	160	286	376	263	268	253	261	179	105	69
%ile	91	96	88	79	83	82	78	80	76	81
$200-399										
mean	171	194	206	220	241	257	270	280	284	299
sd	12	16	18	20	23	25	26	30	32	29
N	252	438	456	469	410	375	249	281	186	119
%ile	93	94	82	76	79	79	77	76	70	75
$199 or less										
mean	166	191	203	222	238	258	265	285	284	299
sd	11	15	17	20	26	24	27	28	25	30
N	130	163	219	204	220	186	137	122	74	45
%ile	87	91	78	79	76	80	73	80	70	75
Range	5	4	8	7	7	7	13	9	14	8

Family Income

There is a significant difference in the achievement of home school students based on family income. As shown in Table 3.10, students in higher income families consistently have higher mean composite scaled scores *(F for income = 79.1; df=3,9186; p < .01)*. There is also a significant interaction of income and grade *(F =2.6; df=27,9186; p<.01)*. Achievement differences due to income are more pronounced for students in higher grades.

Table 3.10
Composite Scale Score Mean, Standard Deviation and Corresponding Percentile by Family Income and Student Grade

Grade	1	2	3	4	5	6	7	8	9	10
$70,000 or more										
mean	173	196	211	225	247	264	278	292	301	306
Sd	10	15	16	20	23	24	25	28	27	29
N	188	300	370	350	296	300	226	202	139	80
%ile	95	96	88	82	85	84	83	85	84	81
$50,000 -69,999										
mean	169	195	209	224	243	261	274	287	293	306
Sd	11	15	17	18	23	24	23	26	29	34
N	165	285	407	352	316	293	239	214	135	109
%ile	91	95	86	81	81	82	80	81	77	81
$35,000 -49,999										
mean	169	193	206	222	241	258	270	281	292	305
sd	12	16	19	21	21	23	26	27	30	30
N	164	266	327	251	269	262	264	212	141	96
%ile	91	93	82	81	79	80	77	81	76	80
$34,999 or less										
mean	167	192	204	218	237	255	262	276	278	297
sd	14	17	17	21	24	28	29	32	30	31
N	149	232	304	245	276	228	178	181	148	66
%ile	89	92	79	74	75	77	70	73	65	74
Range	6	4	7	7	10	9	16	16	13	9

Parent Certification as a Teacher

To determine whether there is a difference in achievement for students in households where at least one parent holds a state issued teaching certificate, we analyzed the data for the 7,607 students with at least one parent that has a college degree. As shown in Table 3.11, the achievement levels across groups are remarkably similar. Controlling for grade and parent education level, there is no significant difference in the achievement levels of home school students whose parents are certified and those that are not *(F for certification=2.9; df=1,7587; n.s.)*.

Table 3.11
Composite Scale Score Mean, Standard Deviation and Corresponding Percentile by Parent Teaching Certificate and Student Grade

Grade	1	2	3	4	5	6	7	8	9	10
At least one certified parent										
Mean	172	196	212	225	245	268	278	289	299	308
sd	11	16	15	20	21	21	24	24	25	31
N	183	293	342	285	290	245	243	208	137	88
%ile	94	96	89	82	83	87	83	83	82	82
Neither parent certified										
Mean	171	195	210	225	246	263	276	291	299	309
sd	12	15	16	19	22	24	25	25	28	27
N	396	688	840	734	661	616	470	412	281	195
%ile	93	95	87	82	84	83	82	84	82	83
Difference	1	1	2	0	-1	5	2	-2	0	-1

Parent Education Levels

The National Assessment of Educational Progress has consistently shown marked differences in the performance levels of students nationwide as a function of parent's educational level. Similar differences appear in the performance levels of home school students. As shown in Table 3.12, at every grade level, children of college graduates out perform children whose parents do not have a college degree (*F=566.4; df=2,9744; p < .01*). There is also a significant interaction between grade and parent education (*F=8.7; df=18,9744; p < .01*), indicating that the effect of parent education is more pronounced in some grades. It is worthy to note that, at every grade level, the mean performance of home school students whose parents do not have a college degree is much higher than the mean performance of students in public schools. Their percentiles are mostly in the 65th to 69th percentile range.

Table 3.12
Composite Scale Score Mean, Standard Deviation and Corresponding Percentile by Parent Education and Student Grade

Grade	1	2	3	4	5	6	7	8	9	10
Both parents have college degrees										
Mean	178	196	212	228	249	268	278	296	306	314
sd	11	15	15	19	21	22	25	22	24	26
N	367	640	706	567	535	501	420	325	206	137
%ile	98	96	89	85	86	87	83	88	87	86
One parent has a college degree										
Mean	172	194	208	222	242	260	275	285	293	304
sd	13	15	16	19	22	24	24	25	28	29
N	212	341	477	451	417	361	293	297	212	147
%ile	94	94	85	79	80	81	81	80	77	79
Neither parent has a college degree										
Mean	161	187	196	212	231	245	260	268	271	288
sd	10	16	17	19	25	25	28	34	27	33
N	121	191	297	255	285	270	233	231	177	104
%ile	79	87	67	66	68	67	69	66	59	67
Range	17	9	16	14	17	23	18	28	35	26

Television Watching

It was pointed out above that home school students spend significantly less time watching television than do the general population of school-age students. Like the nation as a whole, increased amounts of television viewing for home school students is associated with lower achievement test scores. Table 3.13 shows that at every grade level, there is a steady decline in achievement as the amount of television viewing increases *(F for televison viewing =142.5; df=3,9685; p <.01).* The interaction of grade and amount of television viewing is also significant *(F=5.5; df=27,9685; p <.01)*. The effects of television on achievement are more pronounced with students in higher grades.

Table 3.13
Composite Scale Score Mean, Standard Deviation and Corresponding Percentile by Amount of Television Viewing Each Week and Grade

Grade	1	2	3	4	5	6	7	8	9	10
No Television										
mean	166	199	213	227	251	271	281	294	308	307
sd	13	15	15	19	22	24	26	25	27	27
N	81	164	165	161	172	140	117	107	102	64
%ile	87	97	90	84	88	89	86	86	88	81
1 hour or less										
mean	171	196	208	225	245	263	274	288	298	308
sd	12	15	17	20	22	23	25	29	25	29
N	355	554	795	650	586	525	453	369	225	186
%ile	93	96	85	82	83	83	80	82	81	82
2 hours										
mean	169	191	205	219	238	253	268	279	278	299
sd	11	15	18	20	23	26	27	27	31	29
N	186	325	380	333	333	309	237	241	182	92
%ile	91	91	81	75	76	75	75	75	65	75
3 hours or more										
mean	169	187	203	216	233	252	269	275	281	280
sd	11	17	17	20	26	27	28	31	29	35
N	75	121	136	117	135	155	140	130	86	43
%ile	91	87	78	71	70	74	76	72	67	60
Range	5	12	10	11	18	19	13	19	30	28

Summary of Major Findings

Major findings: Demographics

- Home school parents have more formal education than parents in the general population; 88% continued their education beyond high school compared to 50% for the nation as a whole.
- The median income for home school families ($52,000) is significantly higher than that of all families with children ($36,000) in the United States.
- Almost all home school students (98%) are in married couple families. Most home school mothers (77%) do not participate in the labor force; almost all home school fathers (98%) do work.
- Home school students watch much less television than students nationwide; 65% of home school students watch one hour or less per day compared to 25% nationally.
- The median amount of money spent annually on educational

materials is about $400 per home school student.
- The distribution of home school students by grade in grades 1-6 is consistent with that of all school children. Proportionally fewer home school students are enrolled at the high school level.

Major findings: Achievement

- Almost 25% of home school students are enrolled one or more grades above their age-level peers in public and private schools.
- Home school student achievement test scores are exceptionally high. The median scores for every subtest at every grade (typically in the 70th to 80th percentile) are well above those of public and Catholic/Private school students.
- On average, home school students in grades 1 to 4 perform one grade level above their age-level public/private school peers on achievement tests.
- The achievement test score gap between home school students and public/private school students starts to widen in grade 5.
- Students who have been home schooled their entire academic life have higher scholastic achievement test scores than students who have also attended other educational programs.
- There are no meaningful differences in achievement by gender, whether the student is enrolled in a full-service curriculum, or whether a parent holds a state issued teaching certificate.
- There are significant achievement differences among home school students when classified by amount of money spent on education, family income, parent education, and television viewing.

Discussion

Incorporating the largest sample ever used to study home school students and their families, this study is a rich source of information concerning their demographics and achievement. It clearly shows that home school students and their families are a select population. Family income and education levels are well above national averages. The family structure is traditional with married couples as parents, several children, father as bread winner, and a stay-at-home mother. A large percent of home school students have a parent that has held a state-issued teaching certificate. Home school families do not spend a great deal of money on educational materials and tend not to subscribe to pre-packaged full-service curriculum programs.

In spite of the large size of this assessment, there are notable limitations to this study. Foremost, home school students and their families are not a cross-section of the United States population. The act of home schooling distinguishes this group in terms of their exceptionally strong commitment to education and children. There are major demographic differences between home school families and the general United States population. Further, it should be noted that it was not possible within the parameters of this study to evaluate whether this sample is truly representative of the entire population of home school students.

The content of the Riverside tests is another major limitation of this study. While home schools teach the basic skill areas of reading, mathematics, social studies, and science, they do not necessarily follow the same scope, sequence, or emphasis as traditional public and private schools. The primary focus of many home schools is on religious and moral values. Home schools can and do place a greater emphasis on study skills, critical thinking, working independently, and love of learning. Public and private schools usually select the Riverside test due to its close alignment with their curriculum; home schools select the test primarily out of convenience.

We were conservative in our analysis of achievement test results. Even though some 25% of home school students are enrolled in an advanced grade level, we used current grade placement rather than the age appropriate grade placement when determining percentiles and grade equivalents. When looking at test scores, we chose the composite score with mathematics computation, even though mathematics appears to be a weaker subject for older home school students. As a result, we have probably underestimated home school academic performance levels.

Even with our conservative approach, the achievement levels of the home school students in this study are exceptional. Within each grade level and each skill area, the median scores for home school students fell between the 70th and 80th percentile of students nationwide and between the 60th and 70th percentile of Catholic/Private school students. For younger students, this is a one year lead. By the time home school students are in 8th grade, they are four years ahead of their public/private school counterparts.

Our results are consistent with previous studies of the achievement of home school students. A 1990 national home schooling survey of 1,516 families in the United States noted that, on average, home education families have parents with greater formal education, more children, and higher family income (*Home School Court Report*, 1990). Two-parent families were the norm and they were predominantly Christian. The average age of the children was just over eight years--a majority of the children had never attended public or private schools. There were equal numbers of male and female students. On standardized achievement tests, the home-schooled students performed at or above the 80th percentile on national norms in reading, listening, language, math, science, social studies, basic battery, and complete battery scores.

Calvery et.al. (1992) compared the achievement of Arkansas home schooled and public schooled students in grades 4, 7, and 10 using 6 subscales of the MAT-6. Home schooled students scored higher than their counterparts in reading, mathematics, language, total basic battery, science, and social studies at grade 4 and grade 7. They also scored significantly above public school means for grade 10 in reading, mathematics, total basic battery, science, and social studies, but scored significantly lower in language.

Ray (1997) analyzed demographic and achievement data from 5,402 home school students in 1,657 families. While Ray used a different approach to analyze achievement data, he noted exceptionally high average achievement levels and that students with long histories of being home schooled had higher achievement scores.

Home school students did quite well in 1998 on the ACT college entrance examination. They had an average ACT composite score of 22.8 which is .38 standard deviations above the national ACT average of 21.0 (ACT, 1998). This places the average home school student in the 65th percentile of all ACT test takers.

These comparisons between home school students and students nationwide must be interpreted with a great deal of caution. This was not a controlled experiment. Students were not randomly assigned public, private or home schools. As a result, the reported achievement differences between groups do not control for background differences in the home school and general United States population and, more importantly, cannot be attributed to the type of school a child attends. This study does not demonstrate that home schooling is superior to public or private schools. It should not be cited as evidence that our public schools are failing. It does not indicate that children will perform better academically if they are home schooled. The design of this study and the data do not warrant such claims. All the comparisons of home school students with the general population and with the private school population in this report fail to consider a myriad of differences between home school and public school students. We have no information as to what the achievement levels of home school students would be had they been enrolled in public or private schools. This study simply shows that those parents choosing to make a commitment to home schooling are able to provide a very successful academic environment.

Note

This report was supported with a grant from the Home School Legal Defense Association, Purcellville, Virginia. The opinions expressed in this report are those of the author and do not necessarily reflect the positions or policies of the Home School Legal Defense Association.

References

ACT, Inc. (1998). *The 1998 ACT High School Profile Report--National Data*. Iowa City, IA. Available on-line: http://www.act.org/news/98/98data.html

Bruno, Rosaline and Andrea Curry (1997). *Current Population Reports. Population Characteristics: School Enrollment--Social and Economic Characteristics of Students*: October 1995 (update).Available on-line: http://www.census.gov/prod/2/pop/p20/p20-492u.pdf

Calvery, Robert; and Others (1992). The Difference in Achievement between Home Schooled and Public Schooled Students for Grades

Four, Seven, and Ten in Arkansas. Paper presented at the Annual Meeting of the Mid-South Educational Research Association (21st, Knoxville, TN, November 11-13, 1992).

Day, Jennifer and Andrea Curry (1998). *Current Population Survey (CPS) for the Nation*. United States Census Bureau. Available on-line: http://www.census.gov/population/www/socdemo/school.html

Drahozal, Edward (1997). Validity Information for the Iowa Test of Basic Skills (ITBS) and Iowa Test of Educational Development (ITED), Forms K, L, M. Riverside Publishing Company, working draft.

Home School Court Report (Dec. 1990). A Nationwide Study of Home. Available from ERIC Document Reproduction Service ED381725.

Hoover, H.D., N. Hieronymous, D.A. Frisbie, S.B. Dunbar (1996). *Catholic/Private Norms: ITBS*. Itasca: IL: Riverside Publishing Company.

Lines, Patricia (1998). Home schoolers: Estimating Numbers and Growth. Technical paper.

Ray, Brian (1997). Home Education Across the United States. Purcellville, VA: Home School Legal Defense Association. Available on-line:
http://www.hslda.org/media/statsandreports/ray1997/index.stm

Riverside Publishing Company (1994). *Riverside 2000 Integrated Assessment Program*, Technical Summary. Chicago: Riverside Publishing Company.

Scannell, D.P, O.M. Haugh, B.H. Loyd and C.F. Risinger (1996). *Catholic/Private Norms: TAP*. Itasca: IL: Riverside Publishing Company.

Snyder, Thomas and John Wirt (1998). *The Condition of Education*, US Department of Education, National Center for Educational Statistics.

US Census Bureau (1996). *Educational Attainment in the United States* March 1996 (Update). Available on-line:
http://www.census.gov/prod/2/pop/p20/p20-493u.pdf

US Census Bureau (1997a). *Statistical Abstract of the United States, 1997*. Available on-line:
http://www.census.gov/prod/3/97pubs/97statab/labor.pdf

US Census Bureau (1997b). *Statistical Abstract of the United States, 1997*. Available on-line:

http://www.census.gov/prod/3/97pubs/97statab/educ.pdf

US Department of Education (1996). *Youth Indicators*, Washington, DC: National Center for Educational Statistics.

US Department of Education (1997). *National Assessment of Educational Progress, 1996 National Mathematics Results.* Washington, DC: National Center for Education Statistics.

Acknowledgments

This report relied on the creativity and expertise of several individuals. Michael Farris, Esquire, President of the Home School Legal Defense Association conceived the study and secured the cooperation of the author and three contributing organizations - Bob Jones University Testing Service, National Computer Systems, and HSLDA. Earl Hall of HSLDA worked out the numerous logistics and details of working, was the primary architect of the background questionnaire, and provided responses to my thousand questions regarding home schooling. Janet Abbott of BJU provided information about the testing program and the datasets. BJU staff also hand coded the student identification numbers to make corresponding background and achievement information possible. Tom Perry of National Computer Systems handled the logistics of scanning the background questions. Edward Drahozal, Janet Adair and Vesna Plavsic of Riverside Publishing Customer Support were wonderful, knowledgeable resources helping the author understand the ITBS/TAP norming process and data disks. Finally, the author is deeply indebted to H.D. Hoover of the Iowa Testing Program, Gene V Glass of Arizona State University, and Michael Scriven of Claremont Graduate University for their invaluable comments on drafts of this report and for helping to assure appropriate analytical methodology.

About the Author

Lawrence M. Rudner

Email: rudner@ericae.net

Dr. Rudner is with the College of Library and Information Services, University of Maryland, College Park. He has been involved in quantitative analysis for over 30 years, having served as a university professor, a branch chief in the U.S. Department of Education, and a classroom teacher. For the past 12 years, he has been the Director of the ERIC Clearinghouse on Assessment and Evaluation, an information service sponsored by the National Library of Education, U.S. Department of Education which acquires and abstracts articles and manuscripts pertaining to educational assessment, evaluation, and research; builds and maintains on-line databases; publishes articles and books; and provides a wide range of user services. Dr. Rudner holds a Ph.D. in Educational Psychology (1977), an MBA in Finance (1991), and lifetime teaching certificates from two states. His two children attend public school.

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa- DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Maria Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 674 times since March 26, 1999

Education Policy Analysis Archives

Volume 7 Number 9 March 26, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

High School Staff Characteristics and Mathematics Test Results

Mark Fetler
California Department of Education

Abstract

This study investigates the relationship between measures of mathematics teacher skill and student achievement in California high schools. Test scores are analyzed in relation to teacher experience and education and student demographics. The results are consistent with the hypotheses that there is a shortage of qualified mathematics teachers in California and that this shortage is associated with low student scores in mathematics. After controlling for poverty, teacher experience and preparation significantly predict test scores. Short-term strategies to increase the supply of qualified mathematics teachers could include staff development, and recruitment incentives. A long-term strategy addressing root causes of the shortage requires more emphasis on mathematics in high school and undergraduate programs.

Introduction

Debate on how best to teach mathematics has a long history, dating at least to Plato's Greece and the methods illustrated in the Meno. That dialogue presented mathematics instruction as an instance of a general method of teaching based on inquiry. More recent authors frame education as a system, described by indicators of instructional context, processes, and outcomes. (Levin, 1974; Murnane, 1987; Office of Educational Research and Improvement, 1988; Shavelson, McDonnell and Oakes, 1989; and Porter, 1991) Martin (1996) describes an influential framework for the study of mathematics achievement adopted by The Third International Mathematics and Science Study (TIMSS). The TIMSS framework focused on an explanatory system with three factors: what is intended to be taught; what teachers actually do in the classroom; and what students learn.

Curriculum standards and frameworks—"what is intended to be taught"—are teaching tools that can help students to learn, depending on the skill of the teacher. Perhaps because direct measures of teaching skill are difficult to define and obtain, researchers and policymakers use teacher education and experience as plausible proxy measures. Individuals with the same amount of experience and similar teaching credentials can vary in actual skill. Even so, in an aggregate consisting of many teachers in many schools, it is not unreasonable to believe that more highly educated and experienced teachers possess greater skill. Additionally, it is likely that the presence in schools of more educated and experienced teachers is associated with better student achievement.

The departmentalized instruction found in most high schools offers an opportunity to study the linkage between teaching and learning of specific subjects, such as mathematics. Secondary school teachers typically possess single-subject credentials that require proof of specialized subject matter knowledge, and authorize teaching in specific areas. The possession of a valid credential and authorization to teach mathematics is one indicator of teacher education and experience. By contrast, at the elementary level, teachers possess multiple-subject credentials. At the elementary level there is no specific indicator of teacher skill in mathematics instruction.

Mathematics Curriculum and Achievement

Official statements of high school course requirements and curriculum standards permit an inference about the importance of high school mathematics. A national survey of states, conducted by the Council of Chief State School Officers (1998) found that 23 states require more than two credits in math, compared to 13 states in 1989. Forty-two states, including California, have mathematics content standards ready for implementation. Table 1 summarizes the minimum academic coursework expected from California students during a traditional four-year high school career. The California State University System and the University of California publish minimum subject requirements for freshman admission, including electives.

Based on these recommendations, mathematics should be considered second only to English, and should occupy 20 percent or more of a student's curriculum.

Table1
California High School Core Academic Course Requirements

	High School Graduation	California State University	University of California
Mathematics	2	3	3
English	3	4	4
Science	2	1	2
History/ Social Science	3	1	2
Other/ Electives	3	6	4
Total	13	15	15

The National Assessment of Educational Progress (NAEP) (See http://nces.ed.gov/naep/.) has demonstrated influential models of standards-based assessment, and has focused attention on student achievement in mathematics by providing state-by-state summaries of student performance. The NAEP 1992 and 1996 assessments (Reese, et al., 1997) used a framework related to the National Council of Teachers of Mathematics (NCTM) "Curriculum and Evaluation Standards for School Mathematics," originally published in 1989. (See http://www.nctm.org.) These standards include five mathematics strands: number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions. In addition to the five strands, the NAEP assessment examined mathematical abilities (conceptual understanding, procedural knowledge, and problem solving) and mathematical power (reasoning, connections, and communication). Mathematical abilities relate to the knowledge or processes involved in successfully handling tasks. Mathematical power refers to the ability to reason, to communicate, and to make connections of concepts and skills across strands, or from mathematics to other areas. Table 2 displays percentages of students attaining mathematics achievement levels for California, the western region, and the nation. Although California's percentage is one point lower for the most advanced students in 1996, the state is ten points lower for students at or above a basic level.

Table 2
NAEP Grade 8 Percentages of Students at Achievement Levels

	At or Above Advanced	At or Above Proficient	At or Above Basic	Below Basic
1996				
Nation	4	23	61	39
Western Region	3	22	59	41
California	3	17	51	49
1992				
Nation	3	20	56	44
Western Region	3	21	58	42
California	2	16	50	50

Teacher characteristics may explain some of the variation in NAEP scores. Hawkins (1998) used eighth grade data from the 1996 assessment to show that students taught by teachers with an undergraduate or graduate major in mathematics scored higher than students taught by teachers with majors in education or some other field. Moreover, students taught by teachers with certificates in mathematics outperformed students taught by teachers with certificates in other areas. Finally, students of teachers who rated themselves as knowledgeable or very knowledgeable about the NCTM curriculum and evaluation standards scored higher than students whose teachers reported little or no knowledge of the standards.

Achievement of College Entrants

The California Postsecondary Education Commission (1998) estimates that thirty percent of public high school graduates are eligible for freshman admission at the California State University (CSU). Admission requirements in mathematics include three years of college preparatory coursework, normally Algebra I, Algebra II, and Geometry. CSU requires all entering freshmen to demonstrate proficiency in mathematics, either by taking the Entry Level Mathematics (ELM) examination or by presenting proof of adequate performance on an appropriate Advanced Placement, SAT, or ACT mathematics test. Those who cannot demonstrate proficiency must take remedial courses. California public high schools produced 269,071 graduates in 1997 and California State University enrolled 26,781 of them in fall 1998, for a college going rate of ten percent.

Fifty-five percent of these first time freshmen required remedial mathematics instruction. (See http://www.co.calstate.edu/asd/index.html.)

Supply and Preparation of Teachers

Supply and demand for mathematics faculty is one part of a larger system that prepares and employs teachers. Growing student enrollment is driving the demand for qualified teachers. The National Center for Educational Statistics (NCES, 1996) estimates that total K-12 enrollment will grow about 10 percent from 49.8 million in 1994 to 54.6 million by 2006. California public school enrollment will rise over 11 percent from 5.6 million students in 1997 to 6.2 million ten years later. (California Department of Finance, 1998) California public schools employed 270,000 teachers in 1997. (California Department of Education, 1998). Other factors remaining equal, the enrollment growth should create about 30,000 new teaching positions over the next decade. This growth, combined with turnover related to retirement and attrition, and with efforts reduce class size, have resulted in estimates of a need to hire more than 300,000 teachers in California over the next ten years.

Given that teaching skill is associated with student achievement, school districts and policymakers are interested in how teachers are prepared. (Darling-Hammond and Hudson, 1990; National Commission on Teaching and America's Future, 1996; Ashton, 1996; Education Week, 1997) While teaching skill is a goal of preparation, usually a credential only requires an academic degree and coursework. Virtually all public school teachers in the United States have at least a bachelor's degree, and a majority possess an advanced degree. (NCES 1995a) The trend is toward higher levels of education. In 1971, 28 percent of public school teachers possessed a master's, specialist, or doctoral degree. Twenty years later 53 percent of teachers had an advanced degree.

Although high demand for teachers is prompting reforms, California's degree and coursework requirements tend to resemble those of many other states. (National Association of State Directors of Teacher Education and Certification, 1998). Historically, a California preliminary credential required a Bachelor's degree in a subject other than professional education, and a one-year preparation program with training in educational principles and teaching strategies. Those seeking a clear credential fulfill additional course requirements and a year of educationally related study. Career changers with at least a Bachelor's degree and competence in their subject of instruction may work as paid teaching interns while they receive support and training in pedagogy from school districts or universities.

Nationally, schools are filling an increasing proportion of vacancies with inexperienced applicants. (NCES 1995a) From 1988 to 1991 public schools hired more first-time teachers and fewer reentrants or transfers. Teachers who transfer from other schools or return to a school have more experience, but receive higher salaries than first-time teachers. First-time teachers earn less, but are more likely to leave the profession. Teacher retirement and migration into other occupations influence turnover in schools. (NCES 1995b) Nationwide, between 1990-91 and 1991- 92 about 5 percent of teachers left teaching, including retirees. Teachers with less full-time teaching experience were more likely to leave. Smaller schools experience higher teacher attrition. Lower salaries and benefits may be a factor in this relationship. Small schools offer teachers less compensation than larger schools. School with more student poverty have higher turnover than other schools.

Credential requirements restrict access to the teaching profession. One way to meet increased demand is to relax the requirements, reducing the time and cost to become a teacher. For example, when there are too few fully qualified applicants, California school districts use emergency permits to hire individuals who lack some requirements for a credential, usually proof of competence in their subject(s) of instruction or pedagogy. (Hart and Burr, 1996) In recent years emergency permits have become more popular. A risk of this increased popularity is that less well prepared teachers may be less effective in their jobs or more prone to attrition.

States have sought to increase the supply of teachers by setting up alternatives to traditional training programs. Zumwalt (1996) describes alternative certification as easing entry requirements, minimizing preparation needed prior to paid teaching, and emphasizing on-the-job training. Proponents portray these programs as attracting higher-ability, more diverse, experienced people with subject matter majors. (Ashton, 1991; Dill, 1996; Feistritzer, 1994; Haberman, 1992) Zumwalt cautions that it is difficult to generalize about the success of alternative programs. Alternative approaches assume that school staffs have the resources to support unprepared novice teachers. The success of alternative approaches may actually depend on the extent to which novice teachers actually receive needed support and obtain classroom assignments appropriate to their abilities.

Mathematics Teachers

Nationwide, 90 percent of teachers in grades 9 through 12 for whom mathematics is their main assignment report having a major or minor in that subject. (NCES; 1998, 1999) Students at public secondary schools with a higher poverty level or with a higher percentage minority enrollment were more likely to be taught any of the core subjects, including mathematics, by a teacher who had not majored in that subject.

Brunsman (1997) describes California's requirements for a single-subject credential in mathematics and the number of qualified mathematics teachers. In addition to the requirements that apply to all credentials, high school mathematics teachers must demonstrate their competence in the subject either by completing a subject matter program, or they can demonstrate their competence through an examination.

Approved subject matter programs include a core with at least 30 semester units of mathematics coursework that is related to subjects that are commonly taught in departmentalized mathematics classes. This core includes courses in first and second year algebra, geometry, first and second year calculus, number theory, mathematics systems, statistics and probability, discrete mathematics, and the history of mathematics. Programs also include a minimum of 15 semester units of supplemental coursework to provide breadth.

Optionally, a credential candidate can demonstrate subject matter competence in mathematics through examination. California has adopted a standardized subject matter test in mathematics that includes two hours of multiple-choice questions and a two-hour performance assessment. Historically, less than half of the examinees pass the examination.

School districts can assign less than fully qualified teachers to mathematics classes by several methods. An emergency permit requires a Bachelor's degree, passing a basic skills test, and completing a minimum of 18 semester hours or 9 upper division/graduate semester units of course work in mathematics. In order to renew the permit, the teacher must complete six semester units toward earning a credential in mathematics. A limited-assignment emergency permit requires that the teacher have a valid teaching credential in another subject. A waiver requires only that the teacher pass or not ever have taken the mathematics portion of a basic skills test.

Table 3 shows the numbers of single subject credentials, emergency permits, and waivers issued 1993-94 to 1996-97 in mathematics. The number of emergency/waiver teachers in mathematics far outpaces the number of fully qualified new teachers. Over the four year period California granted credentials to 2,689 fully qualified new mathematics teachers, and granted other permits or assignments to 6,339 less well-qualified teachers. Unfortunately, it is not known how many fully qualified teachers actually applied for and accepted jobs in public schools. Virtually every waiver and emergency permit represents an employed teacher. These figures suggest that the supply of fully qualified teachers does not meet current demand. There is a downward trend in the number of fully qualified teachers prepared and possibly hired, and an upward trend in the number of less than fully qualified people actually hired on waivers or permits.

Table 3
First Time or New Type Single Subject Credentials, Emergency Permits, and Waivers in Mathematics

Credentials	**1993-4**	**1994-5**	**1995-6**	**1996-7**
Credentials Via Completed Program	470	475	431	449
Credentials Via Passed Examination	278	218	242	126
Total Credentials	**748**	**693**	**673**	**575**
Total Emergency Permits and Waivers	**1,480**	**1,380**	**1,465**	**1,617**

Student Performance

Some research suggests that teacher skills and ability influence student achievement. Greenwald, Hedges and Lane (1996) reviewed a number of studies of the relationship between school inputs and student outcomes. Some school resources, i.e., teacher ability, teacher education, and teacher experience were strongly related to student achievement. On the other hand, Hanushek's (1996) synthesis of research studies found mixed support for a relationship between school resources and achievement. Although Hanushek did not detect a clear pattern, measures of teacher experience were more consistently related to achievement than measures of teacher education. Ashton (1996) notes that teachers with regular state certification receive higher supervisor ratings and student achievement than teachers who do not meet standards. Teachers without preparation have trouble anticipating and overcoming barriers to student learning, and are likely to hold low expectations for low-income children. Ashton suggests that reducing certification requirements and hiring of teachers who do not meet certification standards, worsens the quality of education of low income children.

Method

The 795 regular California high schools in this study typically serve 1.3 million students per year. About 93 percent of regular high schools offer instruction in grades 9 through 12, although various other grade configurations are represented, most commonly 10-12, or 7-12. These schools reported employing 56,571 full-time equivalent (FTE) teachers in fall 1998, with 14.1 percent of the FTE dedicated to mathematics instruction.

Approximately 600 non-traditional high schools serving about 100,000 students per year were excluded from the study. Generally, non-traditional schools have small enrollments and do not offer the academic curriculum needed to attend California's public universities. Reasons for referral to a non-traditional school could include an unstable home environment, emotional difficulties, pregnancy, etc.

Non-traditional schools diverge from regular schools in serving a population of students with different needs and providing different kinds of services.

The web site for California's Standardized Testing and Reporting (STAR) program (http://star.cde.ca.gov/) provided school average mathematics achievement test scores. The Stanford Achievement Test Series, Ninth Edition, (Stanford 9), was administered to all students in grades 2 through 11 between March 15, 1998 and May 25, 1998. Obtaining direct measures of mathematics skill is probably no easier than obtaining direct measures of teaching skill. Multiple-choice test scores are commonly used as expedient indicators of student skill. The Stanford 9 high school mathematics tests require 45 minutes of examination time and include 48 questions. The content of the tests is oriented towards basic skills and is based on the NCTM framework. Scaled scores were derived using Item Response Theory Rasch model techniques. (Harcourt Brace, 1997). Results for schools testing fewer than 10 students were not available.

Overall, 2.5 percent of students were legally exempted either by parent request or by means of an Individual Education Plan or Section 504 Plan for students with disabilities. Possible effects of selective testing were examined with the help of an estimate of student participation in the assessment. Grade level participation rates were estimated using fall 1998 grade enrollment as the denominator and the number tested as the numerator. Given California's increasing enrollment, this statistic slightly underestimates actual participation statewide.

Teachers with instructional assignments in mathematics were identified from the results of the 1998 Professional Assignment Information Form (PAIF), an annual survey conducted as a part of California Basic Educational Data System (CBEDS). The information requested on the PAIF is required of each certificated staff person, and includes demographics, assignments, and position/credentials. The educational level of teachers with instructional assignments in mathematics was coded as: (1) Doctorate; (2) Master's degree plus 30 or more semester hours; (3) Master's degree; (4) Bachelor's degree plus 30 or more semester hours; (5) Bachelor's degree; and (6) Less than Bachelor's degree. Very few teachers possess less than a Bachelor's degree, so these individuals were aggregated with those who did possess the degree. Years of educational service included service in the current district, other states, and countries, but did not include substitute teaching. School summary statistics for staff with mathematics assignments included the numbers with emergency permits, teaching credentials, and mathematics authorizations. The percent of emergency permits was computed using the headcount of staff with one or more mathematics assignments as denominator and the number of staff with emergency permits as numerator.

AFDC is the percentage of students in the school's attendance area who are enrolled in either public or private schools and who are from families receiving aid. As an indicator of poverty AFDC often correlates with student achievement (White, 1982), and functions in this study as a control variable.

The descriptive and correlational statistics used in this study permit informed speculation about relationships among the phenomena measured by the study variables. Of course, these techniques by themselves do not justify conclusions regarding cause and effect.

Results

Table 4 displays the percent of teachers with one or more mathematics assignments by educational level and possession of an emergency permit. There is a discrepancy between the number of teachers with emergency permits reported by districts on the fall CBEDS census for regular high schools, and a larger number of permits actually issued by the California Commission on Teacher Credentialing (CTC). The CTC number reflects a year- cumulative total for all schools. Differences in the scope and method of data collection likely account for much of the discrepancy.

Table 4
Educational Level and Emergency Permits of Mathematics Teachers

Degree	Emergency Permit	
	Yes	No
Ph.D.	1.3%	1.4%
MA+	5.2%	23.3%
MA	7.7%	16.6%
BA+	26.3%	46.7%
BA or Less	59.6%	12.0%
Total Percent	100.0%	100.0%
Total Head Count	1,009	8,516

The results indicate that 10.5 percent of mathematics teachers in regular high schools have emergency permits. Although some emergency permit teachers have advanced degrees, a majority possessed only a baccalaureate degree. By contrast, the majority of mathematics teachers with credentials completed post baccalaureate work, and about one-fourth of them completed work beyond the masters degree.

Table 5 displays the percent of teachers with one or more mathematics assignments by number of years of service and authorization to teach mathematics. The distributions are bimodal with relatively higher percentages of mathematics teachers with five or less years of experience, consistent with the hypothesis that mathematics teachers tend leave the education profession after several years. At the same time, more than half of all mathematics teachers report ten or more years of teaching experience. Results indicate that one-fourth of those who teach mathematics lack authorization. Although the data do not track teaching assignments over time, about 60 percent of those lacking authorization have ten or more years of experience.

Table 5
Years of Service and Authorization of Mathematics Teachers

Years of Service	**Mathematics Authorization**	
	Yes	**No**
0	0.7%	2.3%
1	7.6%	9.8%
2	5.9%	6.6%
3	5.2%	4.5%
4	5.1%	3.4%
5	5.0%	4.0%
6	4.2%	2.9%
7	3.5%	2.5%
8	3.9%	2.5%
9	3.9%	2.1%
10+	55.3%	59.5%
Total Percent	100.0%	100.0%
Total Head Count	7,228	2,383

Table 6 displays mean test scores, proportion of students participating in the assessment, AFDC, years of experience of mathematics teachers, and education level of mathematics teachers. Grade-level mean test scores and participation rates were weighted by the number of students tested. AFDC and the teacher statistics reflect the entire school and were weighted by total school enrollment. An increasing trend in test scores may indicate improvement in achievement from grade 9 to 11. However, the declining number of students enrolled in higher grades is consistent with student attrition from dropping out. Moreover, the decreasing trend in student participation is consistent with more selective testing in grade 11. Higher scores could be accounted for either by attrition or selective testing.

Table 6
Means of Selected Variables

	School	Grade 9	Grade 10	Grade 11
Test Score	n/a	690	697	703
Participation	n/a	.86	.85	.81
AFDC	15.7%	N/a	n/a	n/a
Years Teaching	14.4	N/a	n/a	n/a
Education Level	3.6	N/a	n/a	n/a
Number Tested	n/a	347,201	313,303	260,933
Number Enrolled	n/a	405,516	370,080	321,896
Number of Schools	n/a	785	794	789

Table 7 displays selected correlations of school mean scaled scores and other variables, weighted by the appropriate grade enrollment. The results appear to be consistent across grades. All correlations are statistically significant, ($p < .001$). Appendix A includes all pairwise correlations of study variables. The empirical correlations probably underestimate the relationships between the study variables for several reasons. The measures of mathematics teacher characteristics are based on a relatively small proportion of all teachers at a school. The student outcome measure, a narrowly defined indicator of mathematics achievement, focuses on one of many areas of study expected of students. More broadly defined measures could produce stronger correlations. Differences in the level of aggregation could also limit the correlations. AFDC and the teacher characteristics are school-wide measures. Test scores and student participation are grade specific. Greater consistency in aggregation, not possible with the available data, could also produce stronger correlations.

The strong relationship often found between poverty and achievement is replicated in this study. AFDC correlates more strongly with test scores than do the other study variables. Correlations with AFDC are largest for ninth grade test scores and smallest for eleventh grade. An investigation of this trend is beyond the scope of this study. However, it is possible that lower achieving students are less likely to be tested in higher grades, possibly the result of attrition or selective testing. If true, the absence of lower achieving students may have resulted in a restriction of range of the variables and lower correlations.

The positive relationship between student participation and test scores is counter-intuitive and seems inconsistent with the hypothesis that lower participation rates are associated with widespread exclusion of lower achieving students. However, school participation rates are negatively related to poverty. Schools with more poverty tend to have lower participation rates and lower test scores. Schools with less poverty tend to have higher participation rates and higher scores. Of course, school characteristics other than poverty could be related to student participation. For example, participation rates might reflect administrative competence. Students at better run schools might have better opportunities to learn and might be more likely to take and do well on tests.

Teaching experience, measured by the average number of years in service, is positively related to test results. Schools with well-prepared teachers tend to have higher mathematics scores, whether preparation is measured as percent of mathematics teachers with emergency permits or as an education level index. To some extent, the effect of teaching experience is mediated by poverty. That is, schools with more poverty tend to have both less well- prepared teachers and lower test scores. One way to assess the influence variables independently is to include all of them in a multiple regression analysis.

Table 7
Correlations for Selected Variables

Variable	9th Grade Test Score	10th Grade Test Score	11th Grade Test Score
Percent AFDC	-0.64	-0.61	-0.59
Percent Participation	0.45	0.48	0.35
Years Teaching	0.24	0.26	0.27
Education Level	-0.24	-0.23	-0.22
Percent Emergencies	-0.39	-0.36	-0.36

Table 8 displays the results of three multiple regression analysis for grades 9, 10, and 11. Achievement test scores were the dependent variables and the analyses were weighted by the number of students tested. The raw weights reflect the variables as originally measured. The beta weights reflect the predictors after scaling to standard deviation units and aid comparisons of the importance of predictors within and across grades.

Table 8
Multiple Regression Analyses by Grade Level

	Grade 9 Weights		Grade 10 Weights		Grade 11 Weights	
	Raw	**Beta**	**Raw**	**Beta**	**Raw**	**Beta**
Intercept	671.1	0	667.5	0	686.3	0
AFDC	-0.6	-10.7	-0.5	-9.3	-0.6	-9.0
Participation	30.7	3.7	39.4	4.5	24.6	2.9
Years Teaching	0.3	1.7	0.4	2.1	0.5	2.4
Percent Emergencies	-27.8	-4.1	-19.8	-3.1	-24.4	-3.2
R-Square	0.50		0.47		0.44	

The three multiple regressions yield similar patterns of results. Student poverty, measured by AFDC, demonstrates the strongest relationship with test scores. Student participation, following the pattern of related simple correlations, is positive related to test scores, even taking poverty into account. The percent of mathematics teachers on emergency permits predicted test scores about as well as student participation. Higher percents of emergencies were associated with lower scores. Finally, the average number of years of teaching experience was positively related to scores. Schools with more experienced mathematics teachers tend to have higher mathematics achievement.

The values of R-square, a measure of how well a combination of the independent variables predicts test scores, appear to trend down as the grade levels increase. This downward trend parallels a similar downward trend in the importance of AFDC as a predictor. One explanation for the trend could be increasing homogeneity of students at higher grade levels. As more disadvantaged students either drop out or find placement in alternative schools, those remaining in regular high schools become more similar socially and demographically. If this hypothesis is true, it could account for some of the increase in test scores in higher grades.

Discussion

The results of this study are consistent with the hypothesis that there is a shortage of qualified mathematics teachers and that this shortage is associated with weak student achievement in mathematics. Student poverty strongly predicts mathematics achievement in this study, as in many others. After factoring out the effects of poverty, teacher experience and preparation are significantly related to achievement.

Several California state policies communicate the importance of learning mathematics. Long standing high school graduation course requirements oblige students to commit a significant amount of time to mathematics instruction. Similar course requirements for college entrance reinforce the message. A state curriculum framework for mathematics appeared in 1985, and the state colleges and universities published a statement of desired competencies in 1982. More recently, the State Board has adopted mathematics curriculum standards for what students are expected to know and be able to do at each grade level. Finally, the current and past statewide assessment programs include mathematics tests. Historically, California policymakers and educators have consistently proclaimed the importance of teaching and learning mathematics. To what extent has the setting of priorities and goals resulted in desired student outcomes?

There are several indications that high school student performance in mathematics does not rise to expectations, for those who are college bound or for others. The 1992 and 1996 NAEP mathematics results are troubling for several reasons. In general, relatively large percentages of students exhibit "below basic" skill levels. Compared to the nation, California has lower percentages of students that are "at or above basic." The NAEP results are consistent the 1998 findings from the STAR assessment program that suggest lagging performance of California high school students on the basis of national norms. Additionally, the California NAEP results do not follow improvements nationwide from 1992 to 1996. Another negative indicator is the finding that over half of 1998 first time freshman at California State University required remedial classes in mathematics.

One explanation for the lower than desired results in mathematics relates to student demographics. Traditionally, student poverty correlates with low achievement. Possibly, disadvantaged students enjoy less support for academic pursuits from their families and peers, and are more focused on meeting needs related to safety and survival. California has a growing number of students whose primary language is not English. These students do not have the same degree of access to the curriculum or assessment as native English speakers. On the other hand, many believe that teaching and learning mathematics depends less on mastery of English than other subjects. Although language skills are important for assessments of writing or reading comprehension, they probably play a lesser role in understanding mathematical notation, solving equations, etc.

Of course, public schools do not control the demographics of their students. Except for those students exhibiting serious disciplinary problems, a public school must serve all who live in its attendance area. There is little that schools can do to change the social and economic circumstances of students. However, the lack of power to alter demographics does not justify complacency towards the education of disadvantaged students. Leverage to improve student outcomes exists at other points in the educational system. Ideally, schools will provide a safe and positive learning environment along with programs and resources to compensate for particular disadvantages. Beyond such compensatory programs, outcomes for disadvantaged students likely depend on sound curriculum standards and quality teaching.

Despite the powerful effect of poverty, the experience and education of mathematics teachers predicts student achievement. Schools with more experienced and more highly educated mathematics teachers tended to have higher achieving students. Schools with higher percentages of teachers on emergency permits tended to have lower achieving students.

Unfortunately, teacher credential information indicates a declining trend in the number of newly-prepared, fully- qualified, high school mathematics teachers, and increases in the number of those who are teaching out of their area or on emergency permits. One reason advanced for these trends is that college students with an interest in mathematics avoid teaching in favor of more lucrative career pathways found in science or engineering. There appears to be a shortage of mathematically able students to meet the overall demand in teaching and other professions. Given growing K-12 enrollments, a strong policy commitment to learning mathematics, and likely growth in technical professions that compete with education, this shortage is likely to persist and grow more severe.

One way to mitigate the effects of this shortage would be to provide training in mathematics to teachers who lack subject matter preparation. One difficulty with such staff development is that the amount of training needed to develop the necessary skills is likely to be great, and that limited staff time and resources will result in long, sustained periods of training. The challenge, in some cases, will be to provide the equivalent of an undergraduate minor spanning multiple courses over a period of years, to a teacher who is already employed full-time. Some under-prepared teachers may not have taken the necessary courses in college because they lacked prerequisite skills from high school. This in-service challenge will be difficult to reconcile with the limited time and resources usually provided for staff development. Given the difficulties, it would be prudent to evaluate the effectiveness of such in-service programs, and to consider other ways of easing the shortage.

Financial incentives might induce more people to take up teaching mathematics. There is abundant anecdotal evidence that higher starting salaries in other fields have drawn people with technical skills away from teaching. One drawback of financial incentives is the potential for inequality and divisiveness that it might create in the teaching profession. An additional issue is whether policy makers could make available sufficient additional funds for an incentive program effective enough to meet the needs of schools.

An alternative long-term strategy to address a shortage would be to require higher levels of mathematical skills of all undergraduate students, possibly by increasing the rigor and number of required lower division mathematics courses, and by requiring more upper division mathematics courses. The general education breadth requirement at the California State University only calls for "a minimum of twelve semester units or eighteen quarter units into the physical universe and its life forms, with some immediate participation in laboratory activity, and into mathematical concepts and quantitative reasoning and their applications." (California State University, 1993) This policy often translates into a requirement for one mathematics course at specific state universities. The general education requirement for students who transfer from a community college only calls for three semester units in "mathematical concepts and quantitative reasoning." Considering the weight given to mathematics in the CSU entrance requirements, the general education mathematics requirement appears inconsequential. Although little has been published on general education requirements, CSU's policies probably resemble those of many other colleges and universities. Strengthening the mathematics requirements could increase the numbers of students who major or minor in the subject, and could help to meet the growing demand for such expertise in teaching and technical professions.

A change in course requirements will face a number of challenges. Some believe that there has been a trend over the last several decades to weaken undergraduate mathematics requirements. One reason sometimes advanced for this trend is that many entering freshmen are not prepared to handle college mathematics. Increasing the rigor and number of required mathematics courses might adversely impact student retention and degree attainment. There may also be difficulty in providing sufficient faculty and resources to support additional requirements in mathematics.

High school student ability in mathematics should be seen as one outcome of a larger system that includes both K-12 schools and higher education. It would be unfortunate if weakened undergraduate requirements are related to poor high school preparation. This pattern could be a symptom of a downward spiral in mathematics literacy in the population. As collegiate requirements are weakened, resources for undergraduate mathematics programs lessen, the mathematics skills of teachers decrease, and the students of these teachers are less well prepared. Expectations of faculty and administrators in high school and college could drift lower, making it more difficult to provide the resources and leadership needed to create and implement high standards. In the short run a pattern of low expectations and low performance is the path of least resistance. Rigorous mathematics courses are not popular with students and are unrewarding for faculty. Easing the requirements provides short term relief. In the long run the path of least resistance results in lowered student ability and decreased capacity to make improvements. Public discontent with school performance will grow unless teaching and learning improve.

Schools and teacher preparation programs need to coordinate their programs more closely in preparing, recruiting, and hiring teachers. One basis for cooperation would be to set policy goals at both the state and local levels to eliminate the use of less than fully qualified teachers within a given time frame, for example, within five years. At the state level it would be useful to reduce the options for hiring less than fully qualified teachers and simplify the procedures for obtaining an authorization to teach mathematics. Subject matter preparation programs are approved partly on the basis of course titles and descriptions. Unfortunately, titles and descriptions permit considerable latitude in the rigor of such programs and there is little assurance as to the skills that prospective teachers actually develop. Given the apparent shortage of mathematically inclined undergraduates there may be an incentive to lessen the rigor of preparation programs in order to keep the "pipeline" full and meet school district demands. One way to cope with variation in rigor is to establish a uniform assessment of subject matter knowledge needed to teach high school mathematics.

The effectiveness of setting goals is reduced and the implementation of well-designed programs is undermined without timely and accurate data that describe how faithfully the programs are implemented and the extent to which outcomes are attained. In particular, monitoring of the supply and demand of teachers is severely hampered when information about credentials, teaching assignments, and employment is scattered across separate agencies or administrative units and is not easily linked. Although one state agency tracks credentials, it does not know how many credential holders are employed in public schools, or elsewhere. Another state agency conducts an annual staff census of teaching assignments, but lacks detailed information about credential status and does not track employment of individuals across time or schools. The agency responsible for teacher retirement maintains some employment history information, but does not follow credentials or assignments. Finally, student outcome data is not readily associated with information about teachers. Employment history, credentials, assignment information and student outcomes should be combined to provide more useful information for policy makers and program administrators.

The possibility of a unified data system in order to guide and evaluate education programs raises legitimate concerns about confidentiality and conditions of employment. Reasonable protection for the rights of individual teachers should be built into any such system. Balancing these concerns for privacy is the need to design, implement, and evaluate high quality programs that work for students. Beyond a responsibility to spend public money wisely there is a moral obligation to prepare students well for success in work and higher education. The use of timely and relevant information is one way to improve the odds for success.

References

Ashton, P. (Ed.) (1991). Alternative approaches to teacher education. *Journal of Teacher Education*, 42 (2), 82.

Ashton, P. (1996). Improving the preparation of teachers. *Educational Researcher*. Vol. 25, No. 9, pp. 21-22.

Brunsman, B. (1997). Recruitment and Preparation of Teachers for Mathematics Instruction: Issues of Quality and Quantity in California. Sacramento: California Commission on Teacher Credentialing. (http://www.ctc.ca.gov)

California Department of Education. (1998). Public school summary statistics: 1997-98. Sacramento: author. http://www.cde.ca.gov/ftpbranch/sbsdiv/demographic s/reports/

California Department of Finance. (1998). California public K-12 projections by ethnicity: 1998 series. Sacramento: Author. http://www.dof.ca.gov/html/Demograp/k12ethtb.htm

California Postsecondary Education Commission. (1996). Performance Indicators of Higher Education. Sacramento: Author. (http://www.cpec.ca.gov)

California Postsecondary Education Commission. (1998). Eligibility of California's1996 High School Graduates for Admission to the State's Public Universities. Sacramento: Author. (http://www.cpec.ca.gov)

California State University. (1993). General education breadth requirements: Executive Order 595. Long Beach, CA: Author. (http://www.calstate.edu/)

Council of Chief State School Officers. (December, 1998). Key State Education Policies on K-12 Education. Washington, DC: Author. (http://www.ccsso.org)

Darling-Hammond, L., and Hudson, L. (1990). Precollege science and mathematics teachers: Supply, Demand, and Quality. In Review of Research in Education (Vol. 16). Washington, DC: American Educational Research Association. (http://www.aera.net)

Dill, V. (1996). Alternative teacher certification. In J. Sikula (Ed.), *Handbook of Research on Teacher Education* (2nd edition, pp. 932-957). New York: Macmillan.

Education Week. (1997). Quality Counts: A Report Card on the Condition of Public Education in the 50 States. Author, Vol. XVI, January 22, 1997. (http://www.edweek.com/)

Feistritzer. C. (1994). The evolution of alternative teacher certification. *The Educational Forum*. 58 (2), 132-138.

Greenwald, R., Hedges, L., and Laine, R. (1996). The effect of school resources on student achievement. *Review of Educational Research*. Vol. 66, No. 3, pp. 361-396. (http://www.aera.net)

Haberman, M. 1992). Alternative certification: Can the real problems of urban education be resolved by traditional teacher education? *Teacher Education and Practice*. 8 (1), 13-27.

Hanushek, E. (1996). A more complete picture of school resource policies. *Review of Educational Research*. Vol. 66, No. 3, pp. 397-409. (http://www.aera.net)

Harcourt Brace. (1997). Stanford Achievement Test Series Ninth Edition Technical Data Report. San Antonio: author.

(http://www.startest.com/)

Hart, G. And Burr, S. (1996). A State of Emergency ... In a State of Emergency Teachers. Sacramento: California State University Institute for Education Reform. (http://www.csus.edu/ier/)

Hawkins, E., Stancavage, F., and Dossey, J. (1998). School Policies and Practices Affecting Instruction in Mathematics: Findings from the National Assessment of Educational Progress (NCES 98-495). Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement.

Levin, H. (1974). A conceptual framework for accountability in education. *School Review*, 82, (3). pp. 363-392.

Martin, M.O. (1996) Third International Mathematics and Science Study: An Overview. In M. O. Martin and D. L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.

Murnane, R. (1987). Improving education indicators and economic indicators: The same problems? *Educational Evaluation and Policy Analysis*, 9, (2), pp. 101- 116.

National Association of State Directors of Teacher Education and Certification (1998). Manual on Certification and Preparation of Educational Personnel in the United States. Dubuque, Iowa: Kendall Hunt. (http://www2.nasdtec.org/nasdtec/)

National Commission on Teaching and America's Future (1996*). What Matters Most: Teaching for America's Future*. New York, Teachers College, Columbia University.

National Center for Education Statistics. (1995a). America's Teachers Ten Years After "A Nation at Risk." Washington, D.C.: U. S. Department of Education Office of Educational Research and Improvement, NCES 95-766. (http://nces.ed.gov)

National Center for Education Statistics. (1995b). Which Types of Schools Have the Highest Teacher Turnover? Washington, D.C.: U. S. Department of Education Office of Educational Research and Improvement, NCES 95-778. (http://nces.ed.gov)

National Center for Education Statistics. (1996). Projections of Education Statistics to 2006. Washington, DC: U. S. Department of Education Office of Educational Research and Improvement. (http://nces.ed.gov)

National Center for Education Statistics. (1998). The condition of education: Indicator 58. Education and certification of secondary school teachers. Washington, D.C.: U. S. Department of Education Office of Educational Research and Improvement. (http://nces.ed.gov)

National Center for Education Statistics. (1999). Teacher quality: A report on the preparation and qualifications of public school teachers. Washington, D.C.: U. S. Department of Education Office of Educational Research and Improvement. (http://nces.ed.gov)

National Council of Teachers of Mathematics. (1989). Curriculum and Evaluation Standards for School Mathematics. Reston, VA: National Council of Teachers of Mathematics. (http://www.nctm.org.)

Office of Educational Research and Improvement. (1988). Creating Responsible and Responsive Accountability Systems: Report of the OERI State Accountability Study Group. Washington, DC: U. S. Department of Education. (http://www.ed.gov/offices/OERI/)

Porter, A. (1991). Creating a system of school process indicators. *Educational Evaluation and Policy Analysis*, 13, (1), pp. 13-29. (http://www.aera.net)

Reese C., Miller, K., Mazzeo, J., and Dossey, J. (1997). National Assessment of Educational Progress 1996 Mathematics Report Card for the Nation and the States. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement. (http://www.ed.gov/offices/OERI/)

Shavelson, R., McDonnell, L., and Oakes, J. (Eds.) (1989). Indicators for Monitoring Mathematics and Science Education: A Sourcebook. Santa Monica, CA: The RAND Corporation.

White, K. (1982). The relation between socioeconomic status and achievement. Psychological Bulletin, 91, 461- 481. (http://www.aera.net)

Zumwalt, K. (1996). Simple Answers: Alternative Teacher Certification. *Educational Researcher*, Vol. 25, No. 8, 40-42. (http://www.aera.net)

Note

A significant portion of the research reported here was completed while the author was a staff member of the Califonria Commission on Teacher Credentialing.

About the Author

Mark Fetler

California Department of Education
515 L Street, #270

Sacramento, CA 95814

916/322-0373 (voice) 916/327-3516 (fax)

Email: mfetler@cde.ca.gov

Mark Fetler earned a doctorate in Psychology in 1978 from the University of Colorado, with training in cognition, measurement, and statistics. He currently works for the California Department of Education Special Education Division in the area of state- and district wide assessment. Past assignments include the California Department of Education Assessment Program, the California Community College Chancellor's Office, the California Commission on Teacher Credentialing, the Northwest Regional Education Laboratory, and the Western Interstate Commission on Higher Education. He serves on the boards of various professional and community organizations.

Appendix A
Pairwise Correlations

Pairwise Correlations

Variable	by Variable	Correlation	Count	Signif Prob
Scaled Score - 10	Scaled Score - 9	0.9487	773	0
Scaled Score - 11	Scaled Score - 9	0.9381	770	0
Scaled Score - 11	Scaled Score - 10	0.9629	778	0
Participation - 9	Scaled Score - 9	0.3873	776	0
Participation - 9	Scaled Score - 10	0.3733	774	0
Participation - 9	Scaled Score - 11	0.3695	770	0
Participation - 10	Scaled Score - 9	0.3883	775	0
Participation - 10	Scaled Score - 10	0.3916	782	0
Participation - 10	Scaled Score - 11	0.3897	778	0
Participation - 10	Participation - 9	0.5876	782	0
Participation - 11	Scaled Score - 9	0.2882	773	0
Participation - 11	Scaled Score - 10	0.2942	780	0
Participation - 11	Scaled Score - 11	0.2913	778	0
Participation - 11	Participation - 9	0.5214	779	0
Participation - 11	Participation - 10	0.7383	787	0
AFDC	Scaled Score - 9	-0.6182	771	0
AFDC	Scaled Score - 10	-0.5863	777	0
AFDC	Scaled Score - 11	-0.5837	775	0
AFDC	Participation - 9	-0.3947	777	0
AFDC	Participation - 10	-0.3145	787	0
AFDC	Participation - 11	-0.2791	784	0
Years Teaching	Scaled Score - 9	0.1711	774	0
Years Teaching	Scaled Score - 10	0.2045	780	0
Years Teaching	Scaled Score - 11	0.2204	776	0
Years Teaching	Participation - 9	0.0588	780	0.1009
Years Teaching	Participation - 10	0.0817	790	0.0216
Years Teaching	Participation - 11	0.077	785	0.031
Years Teaching	AFDC	-0.0843	787	0.018
Education Level	Scaled Score - 9	-0.1852	774	0
Education Level	Scaled Score - 10	-0.1901	780	0
Education Level	Scaled Score - 11	-0.2016	776	0
Education Level	Participation - 9	-0.0663	780	0.0642
Education Level	Participation - 10	-0.1225	790	0.0006
Education Level	Participation - 11	-0.1004	785	0.0049
Education Level	AFDC	0.121	787	0.0007
Education Level	Years Teaching	-0.2919	792	0
Percent Emergencies	Scaled Score - 9	-0.3058	774	0
Percent Emergencies	Scaled Score - 10	-0.285	780	0
Percent Emergencies	Scaled Score - 11	-0.3017	776	0
Percent Emergencies	Participation - 9	-0.0953	780	0.0077
Percent Emergencies	Participation - 10	-0.1145	790	0.0013
Percent Emergencies	Participation - 11	-0.0969	785	0.0066
Percent Emergencies	AFDC	0.1945	787	0
Percent Emergencies	Years Teaching	-0.4056	792	0
Percent Emergencies	Education Level	0.2071	792	0

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E.

BEST COPY AVAILABLE

Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Department of Education

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina—Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

BEST COPY AVAILABLE

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa- DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

BEST COPY AVAILABLE

Este artículo ha sido consultado 224 veces desde el 26 de marzo de 1999

Education Policy Analysis Archives

Volume 7 Number 10 26 marzo 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ
85287-0211

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Investigación en educación en América Latina: una continuación del debate

Abdeljalil Akkari
Fribourg University

Soledad Pérez
Geneva University

Resumen
Recibimos la reacción de Narodowski (1999) publicada por *Education Policy Analysis Archives* 7(2) como una búsqueda de un debate fructuoso sobre la educación en América Latina. En primer lugar, queremos llamar la atención, sobre el hecho que nuestro primer artículo no tiene un carácter exhaustivo, dado que en ningún momento pretendemos presentar todos los aspectos de la investigación educativa sobre América Latina, trabajo que necesitaría varios equipos de investigación y numerosas obras. Nuestro objetivo, modesto por cierto, es realizar un acercamiento a la evolución de la investigación educativa y lanzar el debate sobre los temas que nos parecen prioritarios en la región.

Abstract
We consider Narodowski's reaction (1999) published by *Education Policy Analysis Archives* as Volume 7 Number 2 to be an invitation to a fruitful debate on education in Latin America. We want first to call attention to the fact that our original article—EPAA 6(7)—was not exhaustive; we never attempted to present all aspects of educational research in Latin America, an effort that would require the attention of several teams of researchers and numerous written works. Our objective, modest as it is, is to take a closer look at the evolution of educational research and to launch the debate on those aspects that seem to us high-priority for the region.

Deseando animar aún más el debate iniciado por la revista *Education Policy Analysis Archives*, quisiéramos desarrollar algunos observaciones adicionales.

Identidades latinoamericanas?

La primera crítica que nos hace Narodowski (1999) se refiere a nuestra "simplificación de la identidad de América Latina," la cual sin embargo cuenta con más de 20 países independientes. Nos parece claro que esta identidad latinoamericana es múltiple y en evolución constante, además cada país tiene un desarrollo histórico particular. La división del "Estado-Nación" la cual es privilegiada por Narodowski, nos parece, más bien aproximativa al nivel de la educación. Algunas diferencias, como por ejemplo, la taza de escolarización, dentro de un mismo país son a veces mucho más importantes que las diferencias nacionales entre dos países. En el estado mexicano de Chiapas, las cifras de analfabetismo sirven para ilustran esta situación: Casi uno de cada cinco (19%) de los Chiapanecos de 15 a 29 años es analfabeto, frente a uno de cada 20 (5%) de los jóvenes del país de esa edad. En

Chiapas, el 30% de las personas de 16 años y màs (23% de los hombres y 3 % de las mujeres) carecen por completo de instrucción escolar (CEPAL, 1995).

En Brasil, el ingreso familiar y la ubicación de la vivienda son factores de fuerte influencia sobre la taza de escolarización. En la zonas urbanas, la taza de escolarización de los niños de 10-14 anos llega a 90% mientras que en la áreas rurales esta taza llega a 70% (IBGE, 1994). La falta de escuelas y la incorporación precoz al sistema de producción agrícola explica una parte de esta diferencia.

Desigualdades regionales, entre hombres y mujeres, entre zona rural y urbana, entre población mestiza y población indígena, nos muestran que la identidad de la realidad educativa es problemática mismo dentro un solo país.

Sin embargo, nos parece que existe una herencia histórica común, como también la similitud del proceso de modernización económico-político . Existe sobre todo un modelo de desarrollo dependiente triunfante en el conjunto del continente latinoamericano (con excepción de Cuba), que nos ha empujado a trazar las principales tendencias de la investigación educativa de la región, a poner en evidencia el rol de los diferentes actores y sobre todo intentar de analizar los principales modelos teóricos que han surgido en la región. Existen también dos lenguas internacionales comunes (español y portugués) que favorecen la circulación de las ideas pedagógicas y la colaboración entre los investigadores de los diferentes países.

Pedagogía de los oprimidos

Cuando Narodowski se refiere a un estudio de Feldberg (1995) sobre el sistema educativo Argentino durante el periodo 1976- 1994, señalando que ninguno de los 284 estudios recopilados no se ubicarían en la tradición fundada por Paulo Freire. La explicación de esta situación es evidente. En el periodo mencionado, Argentina está controlada por la dictadura militar, por lo tanto, las ideas de este pensador, que justamente insistían sobre el carácter eminentemente político de todo acto educativo, eran consideradas subversivas como las del Che Guevara. En consecuencia es normal, no encontrar las ideas de Freire representadas en los estudios educativos Argentinos de esta época. Para Freire (1976), los interrogantes más importantes en pedagogía no son estrictamente pedagógicos pero sí políticos. Ademas, el estudio mencionado por Narodowski aparentemente trata sobre todo sobre la educación formal, mientras que el aporte teórico de Freire se refería en la época, más bien a la educación popular, a la alfabetización de los oprimidos, a los marginales, a los excluidos.

Freire lanza a la discusión sus tesis sobre las relaciones entre educación y política, en las cuales prima una lógica de articulación; no se reduce a la educación ni a la economía ni a la política, y al mismo tiempo rechaza la neutralidad de los actos educativos. Los sentidos políticos son constituyentes de la trama educacional, y los procesos educativos producen vínculos políticos particulares. El papel de la conciencia en el proceso de liberación social es muy importante en las ideas de Freire (Puiggros, 1994, p. 446). Y por terminar sobre Freire, podemos decir que todos los intelectuales latinoamericanos saben su aporte en varias experiencias de alfabetización en Brasil, Chile, Perú, Ecuador, Bolivia, Venezuela y otros países.

Carencias estructurales

A nuestra manera de ver, la reacción de Narodowski, no contribuye a animar la discusión, dado que no trata los puntos esenciales mencionados en nuestro artículo, como por ejemplo, el rol mayor y creciente del Banco Mundial que se convierte en actor central de la educación en Latinoamérica. Este organismo focaliza su discurso sobre el mejoramiento de la educación, una mejor gestión de los recursos destinados a la educación, y una adecuación entre las personas formadas y el mercado de trabajo. La intervención del Fondo Monetario Internacional provoca también una reducción de los gastos en educación y una agravación de las desigualdades (Reimers, 1991, Arnove y al, 1996).

La intervención del Banco Mundial en la región está acompañada de investigaciones y evaluaciones hechas por "expertos" quienes mencionan pocas veces las carencias estructurales del sistema educativo de la región. Un estudio reciente del UNICEF (1999) muestra que si los países de América Latina realizan una alta taza neta de escolarización, 92% de los ninos en edad de ir a la escuela primaria son escolarizados, unos de los porcentages más altos del tercer mundo, pero solamente 74% de los niños que comienzan la escuela primaria llegan al quinto grado. Este porcentaje supera de solo 7 puntos la cifra del Africa subsahariana, y se ubica 25 puntos por debajo de las cifras de los países industrializados. Evidentemente, estas cifras tienen el mérito de indicar la tendencia general pero hay también que analizarlas por país y por regiones de cada país. Las diferencias entre regiones rurales y urbanas, entre mujeres y hombres y población mestiza y indígena pueden mostrar situaciones mas dramáticas como la deserción escolar.

La reducción de los gastos públicos sobre educación como lo muestra por ejemplo el caso de Costa Rica donde el presupuesto del ministerio de la educación ha pasado de 30.8% del presupuesto total del país en 1978 a 19.1 % en 1986. Estas cifras muestran el camino que queda a recorrer hacia la meta de la equidad social (Fischel, 1996). En esta perspectiva, pretender impulsar una equidad social gracias al sistema actual de educación es una ilusión (Tedesco, 1998). La relación entre educación y equidad social es compleja y dialéctica.

Educación publica: como reconstruir la utopía inicial? Otro punto importante que quisiéramos mencionar se refiere, a lo que podemos llamar, la construcción progresiva de dos redes de enseñanza básica en la mayoría de los países de la región. Una red pública, abandonada que recibe a los niños de las categorías populares, y una red privada destinada a los hijos de las clases media y alta en casi todos los países. La segregación residencial y económica, bastante marcada hace que los niños según su clase social (privilegiada, media o pobre), pasen no solamente por dos procesos de enseñanza, sino también por varios procesos de socialización, completamente diferentes y desconocidos unos del otros. En esta perspectiva, todo el discurso sobre la ciudadanía y la equidad adquieren una gran importancia estratégica. En realidad, que sentido podemos dar al concepto de ciudadano en la medida que los niños latinoamericanos según su pertenencia social viven sistemas de socialización y de escolarización divergentes.

De esta manera, vemos la agravación de las condiciones de funcionamiento de la educación publica , a tal punto que mismos los educadores que trabajan diariamente en las escuelas publicas, optan cada vez mas por enviar a sus hijos en las escuelas privadas. Por lo tanto, sus esfuerzos para mejorar la educación publica se ve así disminuidos en la medidas que esta situación no le toca personalmente. Señalemos a este propósito que históricamente la escuela obligatoria ha acompañando la revolución industrial en Europa al final del siglo diecinueve, su institucionalización ha estado sobre todo unida a consideraciones socio-políticas (Ramírez & Boli, 1987). Como por ejemplo, acompañar y consolidar el Estado-Nación, realizar una homogeneización entre las diferentes regiones, pero también fomentar una cultura democrática, han sido parte del mito fundador de la escuela obligatoria.

Cuando Jules Ferry hizo votar los celebres leyes escolares francesas que imponen la escuela primaria gratuita y obligatoria todo esto no tenia motivo económico. En la época, ocho franceses sobre diez trabajaban en la agricultura, las minas y los obras publicas, por lo tanto ellos no necesitaban saber leer y escribir. Los patrones estaban satisfechos de una mano de obra iletrada y por supuesto se oponían juntos con la iglesia católica a la creación de la escuela republicana. Los patrones veían en un obrero instruido un revolucionario en potencia o peor aun un sindicalista difícil a manejar (Meirieu & Girard, 1998). El proyecto de Jules Ferry era claramente ideológico y político. La instrucción era considerada como un medio para hacer salir al pueblo de la dependencia de las creencias y supersticiones antíguas para llegar a la racionalidad científica y la democratización política.

Mezclar los alumnos de diferentes orígenes étnicos y sociales, al menos durante la escolarización básica, ha sido una característica central de la escuela obligatoria como invención cultural moderna.

Es interesante mencionar la relación actual entre el modelo de desarrollo económico y el modelo educativo en América Latina. El modelo económico neo-liberal triunfante en Europa y Estados Unidos es copiado y aplicado en varios países latinoamericanos, los cuales intentan seguir los pasos del primer mundo, en todo lo que se refiere al modelo económico, pero no en el modelo educativo. Las diferencias son grandes. En América Latina. La educación está fragmentada de manera precoz, entre educación pública y privada con una diferencia obvia de calidad entre ambas. En el primer mundo, la red privada y publica tienen en la majoría de los estados la misma calidad, sobre todo podemos observar que en la enseñanza primaria, la red privada comporta un porcentaje insignificante de alumnos. Las clases media y altas ponen su hijos principalmente en la red publica. Ademas, en

BEST COPY AVAILABLE

países como Francia y España, por ejemplo, donde existen históricamente una red secundaria privada desarrollada, las autoridades públicas ejercen un control efectivo sobre el programa, la calificación de los profesores y la posibilidad de ingreso por todos los grupos sociales.

De cualquier modo, podemos decir que el debate sobre la educación en América Latina no es un debate sobre las mejores soluciones técnicas (pero un combate político y social para controlar los recursos públicos. Estos recursos son captados en la majoría de los países de la región por los interés particulares de una minoría de ciudadanos (Plank, Sobronho & Xavier, 1994).

La educación puede tener un papel fundamental en el desarrollo nacional a condición de realizar un consenso de todos los actores sociales a fin de llegar a un cambio estructural. El valor de las utopías educativas, como la justicia social, la ciudadanía o la igualdad social, no es que estas puedan ser inmediatamente realizadas pero el hecho de que exista en la visión del educador como perspectivas finales sin que ellas sean olvidadas por motivos pedagógicos instrumentales.

Por lo tanto esperamos a través este corto articulo, respuesta al articulo de Narodowski, impulsar un debate permanente con otros colegas, no solamente referente a la problemática educacional, sino también sobre otros aspectos de la realidad social y política de América Latina.

Referencias

Akkari, A. & Pérez, S. (1998). Educational research in Latin America: Review and perspectives. *Education Policy Analysis Archives*. Vol 6, 7. Available on-line : http://epaa.asu.edu

Arnove, R.F., Torres, A., Franz, S. , Morse, K. (1996). A political sociology of education and development in Latin America. The conditioned State, Neoliberalism, and educational policy. *International Journal of Comparative Sociology*, XXXVII, 1-2, 140-157.

CEPAL (1995). Estudios sociodemograficos de pueblos indígenas. *Notas sobre la economia y el desarrollo*, 578, 1-4. Freire, P. (1976). L'alphabétisation et le "rêve possible". *Perspectives*, VI (1), 70-73.

Fischel, A. (1996). Politics and education in Costa Rica La educacion, *La educacion*. vol especial 40 aniversario. Available on-line : http://www.cidi.oas.org/educ40anivFisch.htm.

IBGE (1994). *Brasil em numeros*. Rio-de-Janeiro: Instituto Brasileiro de Geografia y Estatisticas.

Meirieu, P., & Guirard, M. (1998). *L'école ou la guerre civile*. Paris: Plon.

Narodowski, M. (1999). Educational research in Latin America: A reponse to Akkari and Perez. *Education Policy Analysis Archives*. vol

7,2. Available on-line : http://epaa.asu.edu.

Plank, D., Sobrinho, J.A. & Xavier, A.C. (1994). Obstacles to educational reform in Brazil. *La Educación*, 117, 1, 77-87.

Puiggros, A. (1994). Los limites de las alternativas pedagógicas tradicionales. *Revista interamericana de Desarrollo Educativo*, XXXVIII (119), 441-458.

Ramírez, F.O, and Boli, J. (1987). The political construction of mass schooling: European origins and worldwide institutionalization. *Sociology of Education*, vol.60, 2-17.

Tedesco J.C. (1998). *Les réformes éducatives en Amérique Latine face à leurs défis*. Paris: Institut International de Planification de l'Education.

UNICEF (1999). *La situation des enfants dans le monde 1999. Education*. Genève: UNICEF.

About the Authors

Abdeljalil Akkari

Email: Abdeljalil.Akkari@unifr.ch

Dr. Abdeljlil teaches courses in multicultural education and sociology of education at Fribourg University (Switzerland). He developed research in Tunisia and Brasil on the relationships between education and development. He is currently as a visiting professor in the Psychology Department of the University of Maryland Baltimore county.

Soledad Pérez

Geneva University
Switzerland

Email: perezs@fapse.unige.ch

Dr. Soledad Pérez teaches comparative education at the University of Geneva (Switzerland). She is a specialist of Latin American educational systems. She has done research on rural communities in Equator. She collaborated with international organizations on several projects related to education in Latin America, Europe and Africa.

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

BEST COPY AVAILABLE

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA Editorial Board

BEST COPY AVAILABLE

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North
Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 502 times since March 31, 1999

Education Policy Analysis Archives

Volume 7 Number 11 March 31, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

First Year Implementation of the School to Work Opportunities Act Policy: An Effort at Backward Mapping

Arthur M. Recesso
Valdosta State University

Abstract

This study examined the intent of federal policy and the actual implementation within local school districts. Specifically, the focus is on the Federal School to Work Opportunities Act of 1994 and its implementation in 47 school districts in upstate New York as part of a consortium during the 1995-96 school year. The purpose of the study was to determine 1) the extent to which an agreement to participate in a consortium arrangement designed to facilitate the implementation of a specific Federal or state policy resulted in the active implementation efforts by individual consortium members, and 2) how a high school setting where School to Work activities were perceived by local stakeholders as having great specific and important effects differed from a high school setting where School to Work activities were perceived by local stakeholders as having some or no effect. A bottom-up backward mapping policy analysis model was used for the purposes of this study. Local level data was used to create performance, environment, technology implementation, and School to Work implementation profiles of local high schools. Regression and correlation analyses were used to determine the relationship between stakeholder perceptions and local high school characteristics. Results of the study were reported and interpreted with the aim of furthering research and knowledge of policy analysis and the use of local level data to determine the success of policy implementation. This study found that variation between federal intent and local adaptation is explained by characteristics of the high school and perceptions of stakeholders. School to Work policy implementation, perceived by the high school administrator as a stakeholder, varied significantly by high school student performance, environment in which the high school operates, and level of technology implementation in the high school. Results indicated that the backward mapping policy analysis model is effective in determining the actual levels of policy implementation. Backward mapping results in a definitive explanation of the role of the local actor and the use of discretion in decision making. The final analysis as a result of backward mapping goes beyond the measurement of policy objectives being met and explains the meaning of local level participation.

Introduction

This study compared the original intent of a specific educational reform policy to its local adaptation. Specifically, it focused on the implementation of the Federal School to Work Opportunities Act of 1994 in the forty-four public high schools participating in the New York State Mohawk Valley Workforce Preparation Consortium in the 1995-96 school year.

Policy analysis assumes implementation can be assessed by describing the reported activities of the "local integration" and comparing such findings to the original intent of the policy (Elmore 1982, Ingram & Schneider 1990). The difference between original intent and reported local integration or adaptation were interpreted as "discrepancies" in fulfilling the purpose of the particular policy. This study argued a more inclusive approach. This study was designed to examine these discrepancies in local level implementation. To this end data were collected on both the perceptions of the local stakeholders implementing the policy and the characteristics of the schools in which the implementation took place. To meet the stated objectives and to delineate the patterns of local characteristics and stakeholder perceptions of School to Work implementation, a rank ordering of the forty-four school settings that participated in the Mohawk Valley Workforce Preparation Consortium during the 1995-96 school year

Workforce Preparation Consortium during the 1995-96 school year was undertaken. The rank ordering of perceptions of School to Work implementation ranges from sites where perceptions of having "little or no effect" on the on-going activities in the high school to perceptions of "great influence". The rank ordering on the basis of local characteristics (performance, environment, and technology implementation), showed real deficiencies among sites. The result of the rank ordering of districts on these characteristics served as the basis for the analysis of the relationship between local characteristics and perceptions of implementation.

A second phase of the study described the specific organization, curriculum, instruction, and the community actions related to local implementation in four of the high schools under study. Such description focused on (a) the use of local discretion, (b) the intent of the original policy, and (c) the significance of organizational characteristics. These four sites are described in some detail. The schools were selected on the basis of their ability to elaborate cases of little or no effect, great effect, and a mixed or midrange effect of School to Work as discovered in the first phase.

Background

A Nation at Risk (1983) broadcast a message that schools were not preparing students adequately for the workplace. It sharpened the on-going policy debate about the most effective means for preparing students for the workplace and/or college. The United States was portrayed as uncompetitive in the global marketplace. Students were trained for 1940s and 50s workplaces, not for those of the new millenium. Other reports noted employer complaints that too many recent graduates lacked even the most basic reading, writing, and communication skills (U.S. Congress 1997, America's Choice: High Skills/Low Wages 1983, Nation at Risk 1983, SCANS Report 1983). These reports asked for a policy to address the needs of employers and of students as future employees.

The Role of the Federal Government

The Congressional intent of the School to Work Opportunities Act of 1994 was expressed in fourteen parts to provide an overarching guide for States plans for implementing a sustainable School to Work system. It required the involvement of all key stakeholders, especially employers, in planning the system, implementing it, and supporting its continuation. The School to Work Opportunities Act, as written, was intended to be administered in a way that provided for State and local discretion in implementing the system, as long as the efforts addressed local capacity building, minimized program overlap, used scarce resources effectively, established clear goals, and provided the flexibility to meet these goals (School to Work Opportunities Act of 1994, 108 Stat 568, p.6-7).

In the School to Work Opportunities Act, each State was

BEST COPY AVAILABLE

required to plan three general areas. First, Federal government selection of states to receive funding under the Act was based, in part, on how well this system implementation plan built upon past support for School to Work-type initiatives. (STWOA 1994) Each state's proposal was required to explain how components would be linked together within a five year time-frame to establish a statewide School to Work system. Each state's system implementation plan also was required to show timelines outlining the implementation of all components of the system and the relationship of each to other reform efforts. Second the state must explain how all policy components, including those not supported in the past, were to be supported throughout the state. Thirdly, the state agencies named in the Act by the Federal policy makers to be critical partners in School to Work (i.e., Department of Labor and State Education Agency) were required to be named as having an active role within the State's implementation plan. States were to be funded that demonstrated past practice, a commitment to School to Work-type activities, and planned for the future successful statewide implementation of a School to Work system that met Federal guidelines.

In the first year of approving statewide implementation plans, eight states received funds to implement a statewide School to Work system. State governments decided how funds would be distributed, according to flexible Federal guidelines. The state was responsible for evaluating local progress towards full implementation of the system, ensuring that all students were being served and all stakeholders were involved in the process.

New York State

New York was one of the initial eight States to receive implementation funding. It also was the recipient of the largest grant, over eight million dollars in year one, and more than $60 million over five years. The funds were considered "seed money' and "venture capital" for the development of a School to Work System. (STWOA 1994) New York State submitted a proposal for the implementation of a School to Work System over a ten year period, with five years of funding to be provided in part by the Federal Government (New York State Implementation Grant p.1). Commitment to school reform was established through documentation of workforce development prior to the School to Work Opportunities Act of 1994.

Three-and-three-quarter ($3.75) million dollars had been dedicated to Workforce Preparation sites (involving 129 school districts) in the previous year (1993-94). The Governor's School and Business Alliance was created in 1987; twelve regional sites were provided funding by the Governor's Executive Budget. Both of these initiatives had been working towards building the system called for in the STWOA.

New York State's approved proposal provided the federally mandated framework. Skill standards, skill assessments, and skill certificates were to be integrated into performance-based assessment. Teachers, counselors, and administrators were to be prepared for the

BEST COPY AVAILABLE

Teachers, counselors, and administrators were to be prepared for the implementation of new strategies for curriculum, instruction, and assessment. And parents, employers and other community members were required to be involved in the decision making process for determining appropriate preparation for the workplace or college (New York State Implementation Grant p.6-30). Efforts supported by allocated funds were required to provide opportunities for all students.

Local Partnership

In 1995 the State of New York, through the NYSED, selected fifty-two partnerships throughout the State for funding as planning or implementing sites for School to Work initiatives. These partnerships consisted of school districts, employers, higher education institutions, community members and organizations, and localized State agencies (e.g., County Departments of Social Services). Each partnership had to identify how many and what types of activities were taking place already in the schools, workplaces, and school/workplace projects. They were asked to explain how each activity was to be replicated in non- project sites in the local area and how new activities were going to be developed and supported.

Programs were to be designed and supported for all students, K-12+, including at-risk, gifted, and out of school youth. The program and activity design was to be based on the needs of the local population of students. The needs of the local target population, in turn, was to be determined by student performance on statewide testing, dropout rates, attendance rates, discipline problems, college bound rates, and also local unemployment rates. One result was staff development programs created to address ways to improve student performance. For example, the Teacher Job Shadowing Program connected staff with local employers and employees. Teachers visited business sites, spoke with employers and employees, and toured facilities. Subsequently, each teacher designed a classroom lesson incorporating work-based learning. Some participants were teachers of Regents classes. Their students were to apply knowledge gained through the Regents syllabus to an actual workplace problem. The premise was that students would more thoroughly understand of how the classroom knowledge was applicable to actual problems, resulting in better performance on Regents exams. This was an important point because Regents exams, and the rigors of the courses, were often cited as barriers to change.

Only one region, the Mohawk Valley Workforce Preparation Consortium, submitted a proposal on behalf of the entire region that had been originally designated for funding distribution by the statewide advisory council. Most of the other partnerships in the State were connected and funded through a local BOCES. Other initiatives were funded as either one of the big five school districts (Buffalo, New York City, Syracuse, Rochester, and Yonkers), as a group of school districts participating as a non-BOCES consortium, or as specialized magnet schools like those found in New York City.

BEST COPY AVAILABLE

Conceptual Framework

Many policy studies focus on the role of information and analysis in decision making, others on the role of political bargaining and power in policy decisions (Elmore, 1979, Lindblom & Woodhouse 1993, Wildavsky & Pressman 1971). The actions of local actors and their use of discretion in decision making is often a critical component of these studies (McLaughlin 1987). At any given time, individuals at the local level deal with a myriad of issues and policies related to school reform. Mandates and limited resources force local actors to decide which policies will be implemented. The role of the local actor in implementation may be overlooked in forward mapping approaches to policy analysis. Local actor perceptions and their relationship to decision making and implementation deserve attention that typically does not take place in the traditional top-down approach. If it is agreed that the policy making process yields unintended results (including the failure to implement), then it is desirable to utilize other methods that may aid in determining predictors of success. Important in this approach is the ability to discover forces influencing policy acceptance or integration (Lindblom & Woodhouse, 1993, p.1-32).

Literature Review and Conceptual Framework

Policy Analysis: Forward Mapping v. Backward Mapping

Policy development, implementation, and analysis are extremely complex processes. "As a 'science' policy analysis is concerned with predicting cause and effect relationships implicit in policies" (Elmore 1987 p.174). A traditional approach to policy analysis is the top-down process of forward mapping. Through the traditional lens, the view is of a relationship between policy creator and policy implementation and outcomes. Richard Elmore suggests an approach to policy analysis that goes beyond the traditional top-down analyses used in many studies. Elmore's interest is not limited to determining success or failure based on the measurement of objectives, but seeks to understand why and under what conditions policies are adopted locally. Elmore introduces another vantage-point from which policy is viewed from the locality and by grassroots stakeholders. Both views tell parts of a complex story. Thus a case can be made for studying policy with both the traditional forward lens (from design to local implementation) and a backward lens (from implementation to original intent).

Forward Mapping

The most widely recognized type of analysis is the "top down" approach. This begins with the objectives and goals of the policy and

BEST COPY AVAILABLE

works its way towards the outcomes to be measured in determining success. This process often begins with legislation that outlines the congressional intent of the policy and determines success or failure by measuring reported outcomes in terms of the original objectives.

Supporters of the top down approach assume the clear delineation of goals from policy makers will lead to more effective support, more effective implementation, and greater success of policies in addressing problems. But many authors have reservations about the policy maker's power to affect local implementation process (Ingram & Schneider 1990, McLaughlin 1987). The top down approach assumes the policy makers have control, that they can affect the implementation process, and that their decisions have some bearing on local actors. Authors such as McLaughlin, Ingram, Schneider, and Sabatier fault the top-down approach for not considering the discretion of the local actor as an important component of policy analysis. Ingram and Schneider (1990) cite the work of Elmore for its emphasis on the "organizational structure and personal factors" (e.g., perceptions and discretion) in determining the actual policy impacts and in explaining their success or failure. McLaughlin and others indicate that a need to understand that many factors are involved in the success or failure of a policy, and many of them involve the actions of the "street level bureaucrat" (McLaughlin 1987, Wildavsky & Pressman 1971).

Pressman and Wildavsky's study of the Oakland Project in the 1970's is heralded as a pioneering study of implementation. Their assessment of the Federal implementation of employment programs for severely disadvantaged adults raised questions about why the programs failed. The Federal Government had established very clear goals for the public works projects, the Oakland government officials supported the initiatives, private companies wanted the projects, and the people of Oakland wanted the projects to be a success. Yet the policy failed for reasons that were beyond the control of the policymaker (Wildavsky p.87-124). In their final analysis, Pressman and Wildavsky determined that the relationship between the number of transactions required to implement the policy and the perception that the implementation would be successful were inversely related (Wildavsky 1971). This eventually lead to the understanding that policy decisions are not self- executing (McLaughlin 1987). The policy is created and exists for the local actor to follow as a guide for implementation. The local actor has perceptions of the policy. Those perceptions have a relationship to decision making. Decision-making results in implementing or non- implementing. Therefore, the notion that clearer goals or objectives should have any impact is thwarted and the stance in support of hierarchical procedures is challenged. Pressman and Wildavsky used a forward mapping approach and did not clearly determine the cause of the policy's failure. Their approach does not consider the local actor, the characteristics of the organizations, or the local environment beyond the obvious unemployment rate. Their final analysis does mention the need for a process that takes into account each one of these items. Backward mapping does consider internal and external environment of the local environment and local discretion. And backward mapping completes the analysis by looking back at what was originally intended.

There is no single rule for creating and implementing new policies that will be automatically effective at the local level. Pressman and Wildavsky's study demonstrated that even the best of intentions and the clearest of goals can prove to be ineffective in achieving desired or expected results (Ingram and Schneider,1990, p.69). Still, many analysts believe it is the policy maker and the clarity of the goals and objectives of policy that make the difference at the point of implementation. Ingram and Schneider (1990, p.69) argue that if policy makers framed better statutes, implementation would improve. Expanding on the work of others such as Sabatier, Pressman, and Wildavsky, this theory follows the belief that implementation itself is hierarchical in nature. The focus is on a sequence of events rather than a determination of actual influence or impact (Elmore, Studying Implementation, p.24).

Different processes are required if existing policy analysis approaches are not successful in determining the actual effects of policies (Lindblom & Woodhouse 1993). Often these different, even non-conventional, processes must go beyond the lens of the policymaker. Even though a state or federal government official may play a pivotal role, local influences have an effect on the implementation of the policy. Knowledge of this aspect of implementation may contribute greatly to improving future policy development. Providing data that policy implementation is not necessarily made up of leaders (policy makers) and followers (policy implementors) would be an important step towards establishing policy implementation as an interactive process (Lindblom & Woodhouse p. 4-6). Just as important as following the conventional government process is the investigation into the influences that distort these processes (Lindblom & Woodhouse p.11-12).

Backward Mapping

Backward mapping is a form of policy analysis that acknowledges that there may be variables that exist outside the treatment program, and that some of these variables may have something to do with the organization's characteristics (Elmore 1979, p.601-616). Considering the organizational characteristics of an organization (the school district in this study) could help to explain why policies fail or succeed. While policy makers define the arena in which the implementation process will take place, and determine the actors and the roles each may play, determine the level of resources to be allocated, the impact of environmental influences on implementation cannot always be foreseen (Pressman and Wildavsky p.174, 126). These include organizational, political, and technical conditions (Elmore, Backward Mapping, p. 603).

Backward mapping challenges some of the most basic assumptions of the top down or traditional approach. It does not assume that all organizations are the same. Each has varying environments (e.g., local unemployment, student performance) that affect its decision making related to addressing local problems. The approach does not assume that all organizations are even interested in approach does not assume that all organizations are even interested in implementation. The backward lens places value on the role of the local actor and on the local organization that is focusing on resolution of the problem. The process determines if the school district does not identify itself as having the problem, or has the problem but is not concerned about it. In either event, backward mapping sets the stage for further analysis to determine why either case occurred.

Elmore provides the analysis of the Youth Employment Policy as an example of the shortfalls of using the top down approach (Elmore, 1988, p.29-33). The analysis process begins with the goals and objectives of the policy and summarizes the level of success of the policy by measuring perceived attainment of goals. Similar to the Oakland Project, the results of this analysis may raise concerns that there may be variables outside the process that have a direct impact on the success of the policy (Pressman & Wildavsky, p. 191). However, these other variables go undetected without the use of another approach. McLaughlin adds support for the need of added "bottom up" approach and calls this the micro-level of policy analysis (McLaughlin, Learning from Experience, p.177-178). That is, the focus is on the local actor and the utilization of discretion, acknowledging that the policy is not the only determinant affecting the actions of people at the focal point of the problem (Elmore, Studying Implementation, p.21).

The logic of the analysis used by Wildavsky and Pressman in Oakland and Elmore's assessment of Youth Employment policy indicate the need for further study. It is clear that policy analysis involves not only the measurement of outcomes based on policy goals, but also an understanding of the environment in which the problem resides and within which the implementation takes place. At the local level where decisions are made, the discretion of local actors come into play. It is also where organizational characteristics may have some role in determining the success or failure of the policy. The variation of organizational characteristics, local market conditions, local preferences and prior commitments, and other perceived impacts on the organization will help to determine the relationship between the policy itself, the nature of implementation in the school district, and the actual success of failure of the reform effort.

Design of the Study

The purpose of this study was to compare the original intent of a specific educational reform policy to its local adaptation. The focus was on the implementation of the Federal School to Work Opportunities Act of 1994 in the 44 public high schools participating in the Mohawk Valley Workforce Preparation Consortium during the 1995-96 school year.

Research Question One

To what extent did agreement to participate in a consortium arrangement designed to facilitate the implementation of state or federal policy result in the active implementation efforts by individual consortium members?

BEST COPY AVAILABLE

Research question one, and its subparts, were developed based on the work of several authors such as Pressman and Wildavsky (1971), who raised questions about the meaning of participation and the relationship to implementation. The question also addressed Elmore's (1983,1988) backward mapping model that explains the relationship between the organization's characteristics and implementation. The political bargaining, organizational process, and rational decision making models also partly contributed to the development of research question one. The work of Simon (1993), Allison (1991), Gieseck (1995), and others delineate the need to understand why and under what conditions decisions had been made while implementing the policy.

Research Question Two

How did a high school setting where School to Work activities were perceived by local stakeholders as having great and specific important effects differ from a high school setting where School to Work activities were perceived by local stakeholders as having some or no effect?

Similar to research question one, research question two and its subparts were based on the backward mapping model of Richard Elmore (1988). However, unlike research question one, this question more directly addressed the relationship between local characteristics and implementation of the policy. Research question two was based on the work of Elmore (1988), McLaughlin (1987), Ingram and Schneider (1990), and Lindblom and Woodhouse (1993), all of whom addressed or identified the need to research the differences between local organizations and the relationship of those differences to the level of implementation. In addition, it was within the work of Pressman and Wildavsky (1971) that the question of the relationship between local level characteristics and implementation arose. The research of Rogers (1963, 1988, and 1995) and Tornatzky and Klein (1982) addressed the issue of innovations and their implementation at the local level. Their attention to the organization's involvement in multiple innovations and the relationship between organizational characteristics and implementation of innovations was also a basis for research question two.

Methodology

Forty-seven school districts entered into a partnership to implement a Federal School to Work policy. The question of the match between the federal policy intent and the actual "grassroots" adaptation rested with the level of specific and important changes that can be attributed to four characteristics of the school. School to Work implementation, student performance, environment, and technology implementation as four characteristics of a high school were addressed using a backward mapping policy analysis model.

The characteristics of the schools, to be discussed for the purposes of this study as "profiles," were developed for this study to delineate information for district comparisons, predict changes, profile system strengths and weaknesses, and identify policy improvement strategies (Oakes, 1986). To provide information about the high schools, the profile had to have a reference point, a measure by which to judge the indices used to compare the high schools. The reference points by which to measure the indices were the substantive and technical criteria (Cooley, 1992). Each profile did provide at least one of the following recommended substantive criteria: information that described student performance in terms of achieving outcomes, information that described the central features of the high school, or information that was policy relevant (Oakes, 1986). In addition, each profile was developed having the following technical characteristics: it measured features that can be found in all schools being studied, the data collected can be traced over time, the profiles can be understood by practitioners and policymakers, and the profile indices were generally accepted in educational policy analysis (Cooley, 1992; George, 1993; Oakes, 1986; Windham, 1990). The purpose of developing these profiles was to delineate the local conditions of the high schools implementing federal policy, provide a common means for comparing those schools, and provide a final analysis that improved policy decision making.

The meaning of participant and non-participant as defined by implementing and non-implementing schools was explained using the School to Work implementation profile. The profile described to what extent federal policies were adapted. The School to Work implementation profile also described to what extent the high school implemented school-based learning, work-based learning, and connecting activities components of the School to Work system. The components were required by the federal STWOA for schools using funds to implementing School to Work.

The following data were collected as indices of the level of School to Work implementation achieved in each district: if the school described itself as being involved in School to Work during the 1995-96 school year; the priority of staff development overall and specifically connected to school based learning, work based learning, and connecting activities; the number of partnerships established locally; and the level of commitment to required activities as described in the legislation and regional proposal (e.g., job shadowing, school based learning, tours) as demonstrated by implementation by grade levels. These data were used to compare schools within the consortium. Correlation methods were used to determine the relationship between implementation level, the performance of the school's students, the environment in which the school operates, and the level of technology that has been implemented in the school.

The performance profile of the students in the high school was used to explain the school's educational quality and its effectiveness in meeting desired educational outcomes (Cooley, 1992; George, 1993; Windham, 1990). The New York State Regents exams and diploma are benchmarks by which students and schools were measured to determine academic success. A broad audience of educators and policymakers understand and generally accept the Regents system to be a consistent measure of student achievement (Wiles, 1996). In developing the performance profile, the following data were used: attendance rate, dropout rate, suspension rate, percentage of students graduating with a Regents Diploma; percentage of students enrolled passing the Regents exams in English, Math III, Chemistry, US History and Government; and the percentage of students passing the Occupational Education Proficiency exam. The performance profile also was used to determine a relationship between past performance and present participation rates. The performance profile was designed to measure the school's ability to prepare students to meet expected academic outcomes. This performance measure was compared to the school's ability to meet policy implementation outcomes as described by the School to Work implementation profile. In this analysis, the school's effectiveness in meeting expected outcomes and the level of School to Work policy implemented by that school were tied to the economic viability of the student. Educational outcomes (employment, achievement in subsequent education and training, and admission to further training upon graduation) were all indicators of school effectiveness in meeting educational outcomes (Windham, 1990). These outcomes matched goals of the School to Work policy implemented within these schools (STWOA, 1994). The School to Work policy was developed to better prepared students for the workplace so they could enter the workforce more easily and prepare for better jobs (Jennings, 1995). Jennings found the better the student preparation programs and the better students perform academically, the more likely they are to be employed after high school. This established a link between School to Work and student performance and substantiated the use of performance profile. The performance profile also explained the similarities and/or differences between schools in terms of their effectiveness, and it was used to explain the relationship between student performance and the School to Work implementation profile.

The environment profile was a composite of indices of the external environment and external forces acting upon the high school. These environmental indicators describe the demographic, social, and economic influences of the local communities in which the high schools are located (Bryson, 1988; Cook, 1990). Influences such as types of employers and the number of businesses located in the community were important because School to Work policy makes specific mention of the relationship between successful School to Work implementation and these influences (Grumman, 1994; STWOA, 1994). For example, the school was required to be supporting the work-based learning component. Therefore, employers were required to be involved as partners in School to Work activities. A locality with numerous employers was potentially able to support numerous work-based learning activities.

The availability of resources in the community also described the context in which the schools operate and were represented in the Environment Profile (George, 1993). Levels of per capita income and number of people employed directly influenced school budgets and the funding available to support innovative programs (George, 1993). While School to Work did provide implementation resources to the schools, schools did incur expenses during implementation. Therefore, schools with more resources had the potential to support more School to Work activities.

The environment profile was composed of the following items: percentage of free and reduced lunch (a local wealth indicator), average household income, percentage of people with bachelors degrees, and percentage of youth at risk. These indicators of school and community were then linked to desired outcomes. Each of the profile indicators could be measured for each school district in the School to Work consortium and throughout the state of New York (Oakes, 1986). They capture enduring features that are recognizable to both practitioners and policymakers and generally accepted as being valid and reliable items (State Education Department, 1993; Wiles, 1996).

School to Work took place within the broader context of school reform. A high school that was involved in School to Work was more likely to also be involved in another systemic reform initiative (Grumman, 1994). The School to Work policy was being implemented at the same time as at least one other innovative reform effort in all of the Mohawk Valley Consortium schools. Computer-related technology was also infused in the Mohawk Valley Consortium schools, and its study was completed for most of the same schools during the 1996-97 school year (Mann, 1997). The Mohawk Valley region technology implementation study considered the policy implications for a large-scale commitment to and investment in computer technology. Mann collected detailed information from superintendents, high school administrators, teachers, and students. Data was compiled and manipulated in an effort to determine the impact of technology on student performance (Mann, 1997, p.2).

Technology was used in this study as a means of profiling and comparing schools for several reasons. The data set collected by Mann covered a majority of the same school districts as this study. Similarly, Mann's survey results also relate to level of participation, level of staff commitment, participation in staff development, and the intent of technology integration. And Mann's study sought to explain, in part, how stakeholders perceived the effect, importance, and level of technology implementation. Like this study on School to Work, the Mann study took the approach of addressing implementation at the grassroots level and compared it to the intent of policies in place. Therefore, the fourth and final characteristic was the technology implementation profile. Technology implementation in this study was defined as the level of computer-related technology that had been implemented for instructional use (Mann, 1997). As innovations, School to Work and technology have some distinctive similar qualities that are called the critical mass aspect (Rogers, 1988). Neither innovation reached its potential for utility unless other people or organizations implemented the idea. Technology can operate stand-alone, but its' uses increased as it was connected to other people, the Internet, and networks of software and information. Similarly, School to Work could have been implemented with just one employer or in just one school. But it reached more students and a wider variety of students' needs with multiple participants and supporters. A collaborative effort was required to establish a system that serves all students.

School to Work and technology are similar in the critical mass aspect in many other ways. In order for other organizations and people, other than those directly involved in initial implementation, need to be networked into the continuation of implementation and programs being supported in order to replicate new ideas and see that the innovation works (Rogers, 1988). In School to Work, the more employers and teachers that were involved, the more students that were being served by its initiatives. Similarly, the more technology that was made available to teachers and students, the more access students have leading to a higher level of integration into the learning episode (Mann, 1997). Therefore, the technology profile was an indicator of the extent to which the school had implemented another innovation. Determining the relationship between the technology implementation and School to Work implementation profiles explained the variance attributed to the level of technology implementation. The technology profile was also combined with the environment and performance profile to explain the level of variance in School to Work implementation.

The specific items used to develop the technology implementation profile were as follows: perception of whether the technology importance was linked to improved curriculum or to changes in teaching; and the number of computers installed, networked, accessible to teachers, accessible to students, and the number of Internet accounts in the classroom. These were exactly the same items selected, in part, from the Mann study (1997) and replicated in eleven other high schools that had not participated in that study. (Note 1)

All four profiles of the high school were completed within the framework of a backward mapping policy analysis that took into consideration the conditions of the high school as it implemented the policy. The analysis of the School to Work policy implementation compared the Act's original intent with local implementation that was supported by the organization. The variance between intent and local adaptation was to be explained by the characteristics of the local high school.

The "mapping" process began at the local level and traced the actions back to the beginning. A backward mapping lens, defined as stakeholder perceptions of specific and important effects and local organizations' characteristics within a community environment, was offered as an alternative to traditional top-down methods (Elmore, 1979; McLaughlin, 1987). Stakeholders were considered as any group or individual who was affected by or who could affect the future of the organization (Bryson, 1990). For the purposes of this study, the focus was on local stakeholders with a vested interest in activities that impacted the school district and the student population. These included superintendents, school district administrators, teachers, employers, students and community members. School district administrators provided leadership and decision making for the schools. It was their duty to see that programs were preparing all students leaving school for college or the workplace. The teachers prepared students along a continuum of learning that culminated in graduation from high school and entrance into higher education of the workforce. Whichever path was taken, the students must have had attained the skills and knowledge necessary to perform in their chosen environment. Community members that provided representation on local partnership committees (e.g., private employers) had an interest in students becoming contributing members of the community. The student's contribution could be as a future employee with a good work ethic or as a good citizen that values a role in community efforts. While the forward map would stop at the program administrator's response to the level of integration, the backward map looked for evidence of change in the organization, such as methods used in the classroom, and the interpretation of integration from the viewpoint of participants and stakeholders.

Measuring the level of specific and important changes is an element of the backward mapping model. Grassroots level investigative procedures were used to determine perceptions of School to Work policy implementation resulting in changes in the high schools of Mohawk Valley Workforce Preparation Consortium during the 1995-96 school year. Interpretation of stakeholder perceptions were interpreted to determine if the federal policy intent was similar or significantly different from local implementation. These perceptions were examined in the context of the on-going high school operations, organizational characteristics of the school district, and the local community.

The development of the performance, environment, technology implementation, and School to Work implementation profiles provided a means for describing groupings of data collected in support of the backward mapping model. It enabled the backward map to factor in the impact of external forces such as the local community, other government agencies, and private employers. Local adaptation of program intent, and the conditions within which it took place, was discussed in terms of how the School to Work System was implemented within the local school district. Representative stakeholders of the groups involved in the implementation process described to what extent they perceive School to Work to have been integrated into the schools.

The level of implementation sought by the federal government was delineated to guide the states and schools in their efforts to implement a School to Work. Full implementation was defined in the School to Work Opportunities Act legislation as having these components: (STWOA, 1994)

- All students are participating in School to Work activities,
- All levels of education are supporting an integrated School to Work System,
- All staff have seamlessly integrated School to Work into their classroom,
- All stakeholder groups have been involved in the process, and
- All initiatives and partnerships are tied together in one cohesive system across the State and country.

Schools were differentiated based on the varying degrees to which they were able to achieve full implementation. These levels of implementation were to be determined by: the number of School to Work initiatives that were implemented grades nine through twelve; the number of staff development related activities that were in place; the number and types of partnerships that have been established with local employers; and the number of school-based, work- based, and connecting activities that were supported. The Performance Profile as an independent variable construct was developed to represent the educational quality of the school and its effectiveness in meeting state educational benchmarks. Schools accumulated points in the form of data items on student performance. The total number of points for each high school established a ranking by which to compare the schools. The comparison discussed the success of the school in meeting NYSED-established benchmarks as a member of the consortium and an implementor of School to Work. Higher profile rankings indicated schools having greater success meeting the statewide student performance standards. The success of the school in meeting these standards and, therefore, attaining a higher ranked profile spoke to the educational effectiveness of the school. The Performance Profile was utilized to determine this independent variable's ability to explain the variance in the dependent variable, School to Work implementation. The variance was measured directly and in combination with the two other independent variables.

The Environment Profile was a second independent variable. This profile was a representation of the external factors effecting the school. Highly ranked schools were operating in a healthier (e.g., more opportunities for kids with more employers) environment than other schools participating in the consortium. The availability of employers and employees, community financial resources, and availability of people to serve as appropriate role models could impact the school's ability to support School to Work activities.

Technology Implementation, a third independent variable, was developed based on the responses from high school principals on the level of computer-related technology and training present in each building. The highest ranked district had the most involvement in implementing technology, demonstrating support and dedication of resources to another reform effort.

Sample Selection

The Mohawk Valley Workforce Preparation Consortium consisted of a population of 47 school districts. Forty- four school districts had full secondary programs, grades 9 through 12 inclusive. This was a study of the population of school districts, as partners in the Mohawk Valley Workforce Preparation Consortium, which support high school-level implementation. Therefore, three districts without a grade 9 through 12 program configuration were not surveyed. One district provided schooling up to the ninth grade and transported students to a nearby district for tenth through twelfth grade. The two other districts did not provide a high school program at all. They provided K-8 and K-6 programming and transported students to other districts for further schooling.

One high school administrator from 44 high schools in the Mohawk Valley Workforce Preparation Consortium was surveyed. High school administrators who were not employed by the school district for a full school year during the first year of implementation (1995-96) were not sent surveys. Out of the 44 possible districts, 33 superintendents and 39 high school administrators were surveyed.

Response Rate

High school administrators returned 24 of the 39 surveys, a 62% rate of return.

Summary of Data Collection

To complete this study using backward mapping, data that describe the characteristics of the local organization, the perceptions of local stakeholders, and the adaptation of intent of the policy at the local level was collected. Three types of data collection sources were used. Original surveys were developed and sent to stakeholder groups. Follow-up interviews with stakeholders provided more in- depth information and explanation for the survey data. Existing data sets were readily available to build the profiles that characterize the high schools' student performance, environment, technology implementation, and School to Work implementation. All three data sources proved to be useful and effective in supporting the analyses to be used in this study using the backward mapping process.

Data Analysis

The data collected were analyzed in stages. In the first stage, the data were compiled into profiles, labeled as independent and dependent variables, and analyzed using regression and correlation analysis. In a second stage, the aggregate data from specific questions posed to a surveyed stakeholder group and specific items from the existing data sets were used for analysis. For example, specific questions in the survey were asked to determine if high school administrators responsible for policy implementation described School to Work as being specific and important. Correlation analysis was used to determine the extent to which the administrators' perceptions of importance were associated with the level of implementation. In a second step, regression analysis was applied to determine the extent to which the level of implementation in high schools varied by the perceptions of the high school administrator with significance. The third stage of analysis involved selecting high schools for follow-up interviews by placing the high schools on a quadrant grid. Based on the school's placement on the grid the school was selected for a follow- up interview. Interviews provided more detailed explanations about perceptions, participation, and implementation.

Compiling the Profiles

The profiles (Performance Profile, Environment Profile, Technology Implementation Profile, and the School to Work Implementation Profile) were developed to serve as a means for explaining how organizational characteristics in terms of student performance, external environment, and technology implementation predicted or were associated with the level of School to Work implementation. The profiles developed in this study were essentially constructs. They served the purpose of grouping many different but related variables into one term for ease of analysis and subsequent discussion. The profiles were used to explain to what extent student performance, the environment in which the school operated, and whether the level of technology implemented had an effect on the adaptation of intent at the local level.

The Performance Profile was developed for each high school participating in the Mohawk Valley Workforce Preparation Consortium to determine if school organizational effectiveness characteristics varied systematically with stakeholder perceptions of specific and important effects of School to Work policy. The Performance Profile was a construct consisting of the following variables: percentage of graduates receiving Regents Diplomas, dropout rate, attendance rate, suspension rate, percentage of enrolled passing the Math III Regents Exam, percentage of enrolled passing the English Regents Exam, percentage of enrolled passing the Chemistry Regents Exam, percentage of enrolled passing the U.S. History and Government Regents Exam, and the percentage of number tested that passed the Occupational Education Proficiency- Introduction to Occupations State Exam.

All of the performance data was collected from the New York State 1995-96 Report Card on the Schools released in February of 1997 (Report Card on the School, 1997). Developing the Performance Profile, the reported figures, as raw data, for each item (e.g., percentage of graduate receiving Regents Diplomas) were summed into a total of performance profile points for each high school (See Table 1). The dropout rate and suspension rates were added to the schools total by subtracting the reported rate (e.g., 5% drop-out rate) from 100% and resulting in a positive figure (e.g., 95% did not drop-out) to be added into the schools total (See Table 1). This process thus penalized schools for higher dropout or suspension rates in either category. None of the variables were weighted because there was no evidence of one item having more importance than another (e.g., Math III Regents Exam compared to the Chemistry Regents Exam). The state reporting does not differentiate the items as being more or less important. The total for each high school was rank ordered. The high school with the highest number of total points was given the top ranking. This resulted in a rank order listing of the highest- performing to lowest-performing high schools in the Mohawk Valley Consortium.

Table 1. Compiling the Performance Profile

ITEM	CALCULATION	RESULT
% Grads receiving Regents Diplomas	Actual %	(0-100%)
Dropout Rate	100% - rate	(0-100%)
Attendance Rate	100% - rate	(0-100%)
Suspension Rate	100% - rate	(0-100%)
% of Enrolled Passing:		
Math III Regents Exam	Actual %	(0-100%)
English Regents Exam	Actual %	(0-100%)
Chemistry Regents Exam	Actual %	(0-100%)
US History and Govern Reg. Exam	Actual %	(0-100%)
% of Number Tested & Passed Occ Ed Intro to Occ State Exam	Actual %	(0-100%)
TOTAL	Sum of figures	0 to 900

To determine if community characteristics varied systematically with stakeholder perceptions of specific and important effects of School to Work policy, the Environment Profile for each high school participating in the Mohawk Valley Consortium was developed. The Environment Profile was a construct of the following variables: percentage eligible to receive free and reduced lunch, percentage of households on public assistance, percentage of residents with bachelors degrees, and the percentage of youth at risk.

All of the environmental data was collected from the United States Census Bureau: 1990 Census Data. The sole exception was the free and reduced lunch eligibility data, which was obtained from the 1995-96 Report Card on the Schools. The data were summed for a total of environment profile points for each high school (See Table 2). The totals for each school were then ranked. A lower total number of points indicated a healthier and more resourceful environment in which to operate. Therefore, schools with lower totals were ranked higher: the highest ranking was number one, the lowest, forty-four. The average household income and the percentage of residents with bachelor's degrees had to be resorted to fit the pattern of the other variables. The percentage of residents with bachelors degrees was subtracted from the highest possible percentage (100%). The result was summed into the total environment profile points (See Table 2). The data items were not weighted because none of the data items carried more importance in characterizing the environment of the high school. Data was available for all of the participating high schools.

Table 2
Compilation of the Environment Profile

ITEM	CALCULATION	RESULT
% Eligible for Free and Reduced Lunch	Actual %	(0-100%)
% of Households on Public Assistance	Actual %	(0-100%)
% of Residents with Bachelors Degrees	100 - %w/ BS	(0-100%)
% of Youth at Risk	Actual %	(0-100%)
TOTAL	Sum of figures	0 to 400

To determine if technology implementation varies systematically with stakeholder perceptions of specific and important effects of School to Work policy, a Technology Profile was developed for each high school participating in the Mohawk Valley Consortium. The Technology Profile consisted of the following variables: number of computers installed in the high school, number of computers with a 486 megahertz or faster processor, number of computers networked, number of teachers with access to computers, level of demand from teachers, level of demand from administration, level of demand from community, importance of improving curriculum, and the importance of improving teaching (See Table 3). Each item was summed into a technology profile points total. Similar to the performance and environment profile, no weighting was used with the technology profile. No item was determined, based on the literature reviewed, to be in need of weighting. All items were treated with equal importance.

The level of technology implementation was used as a predictor of involvement in and commitment to reform efforts. The assessment of each school's integration of computer-related technology was intended to explain the level of technology present in the district, the knowledge and attitude of school district staff towards innovations, and more significantly, how technology was being utilized in the learning environment. The variance between the school's technology profile and the level of implementation of School to Work was discussed to explain any patterns or distinct occurrences.

Technology implementation levels were based on participation in the activities that were supported by the Mohawk Regional Information Center (MORIC). The MORIC is the vehicle through which each district purchased computer- related technology, integration strategies, planning services, network support services, and leveraged State Aid for purchases of equipment and services. High implementors had participated in a variety of activities, such as teacher training, curriculum integration of technology, and administrative use of technology. Low implementing schools had few activities or none at all.

Table 3
Compilation of the Technology Implementation Profile

ITEM	CALCULATION
Number of computers in the high school	Actual Number
Number of computers with 486+ processor	Actual Number
Number of computers networked	Actual Number
Number of teachers with access to comp.	Actual Number
Level of demand from teachers	Survey Number
Level of demand from administration	Survey Number
Level of demand from community	Survey Number
Importance of improving curriculum	Survey Number
Importance of improving teaching	Survey Number
TOTAL	Sum of figures

To determine how School to Work implementation varied systematically with stakeholder perceptions and/or the performance, environment, and technology implementation profile, the School to Work Implementation Profile for each participating high school was developed. A School to Work implementation profile was created based on the survey results of the high school administrators.

The High School Administrators' School to Work implementation profile was a construct of the high school administrators' perceptions of the extent to which School to Work implementation took place (See Table 4). It was developed from the survey of high school administrators and used the following variables: provision of multiple learning opportunities (applied academics, career academies, work sites as classroom, work-site simulation, community service projects, internships/apprenticeships, and co-ops/youth-run enterprises), number of grade levels School to Work activities took place (employer visits, job shadowing, simulated work environment, school assignments in the workplace, youth-run enterprise, community service projects, internships, work-linked learning, work-study co-ops), and the level School to Work staff development was supported (active/hands-on learning, multiple learning environments, partnership strategies, development of world-class standards, integrating career development). The high school administrators' total points for School to Work Implementation was summative of all the variables for each high school. The legislation and literature available do not differentiate on the importance of each item. Therefore, no weighting was used in the calculation.

Table 4
High School Administrators' School to Work Implementation

ITEM	RESPONSE RANGE
Provided multiple learning opportunities	
Applied academics	0 to 4
Career academies	0 to 4
Work sites as classrooms	0 to 4
Community service projects	0 to 4
Internships/apprenticeships	0 to 4
Co-ops/youth run enterprise	0 to 4
Number of grade levels STW activities offered	
Employer visits	0 to 4
Job shadowing	0 to 4
Simulation of work environment	0 to 4
School assignments in the workplace	0 to 4
Youth-run enterprise	0 to 4
Community service projects	0 to 4
Internships	0 to 4
Work-linked learning	0 to 4
Work-study co-ops	0 to 4
School to Work Staff Development Supported	
Active/hands-on learning	0 to 10
Multiple learning environment	0 to 10
Partnerships strategies	0 to 10
Developing world-class standards	0 to 10
Integration of career development	0 to 10
TOTAL (sum of response range figures)	0 to 114

Methods Summary

This was not a traditional forward mapping (top-down) analysis. If this study had used the forward mapping approach, the data collection methods would have used the objectives stated in the original policy to determine what data should be collected. This study used the backward mapping (bottom-up) approach. Backward mapping is concerned with the situation in which the policy was being implemented. Therefore, methods different from the top-down approach were used. The data collection methods gathered information about the actual activities at the local level. A final analysis was completed of the perceptions of local stakeholders, the competition for resources, and the environment in which the school was operating to determine their effect on implementation of the policy.

The methods of analysis used in a forward mapping process could have measured the locality's level of implementation by the objectives of the original policy. But that process would not describe the contributing factors or barriers to implementation. Nor would it describe actual implementation and what it meant to a high school. To follow-through on the backward mapping approach, a different method of analysis was used. The perceptions of the local actor as a stakeholder and/or implementor were compared to the original intent of the policy. The extent of variance that can be explained by these perceptions was measured. This analysis of perceptions was in combination with the measurement of variance of School to Work implementation levels attributed to the environment in which the school operates, the performance of its students, and the level to which it has been active in implementing technology.

Results of Data Analysis

Within the context of this section of the article descriptive, regression, and correlation statistics are used to analyze the data collected from the surveys and data sources described. The data include regional and local level data collected from the New York State Education Department (NYSED) and the New York State Department of Labor (NYSDOL); the stakeholder survey responses; and the data collected to build the Environment Profile, Performance Profile, Technology Implementation Profile, and School to Work Implementation Profile.

Findings

High School Administrators: Description of Survey Responses

The high school administrators' information about their school's involvement in School to Work was used to create the second of three subgroups for the School to Work Implementation Profile. The profile was created using the same method as for the superintendents. Each high school received a total number of points based on the high school administrator's reporting of School to Work implementation activities (See Table 4). The distribution of high school for each point range is reported in Table 5.

Table 5
HS Administrator: STW Implementation Points per High School

TOTAL POINTS	NUMBER of SCHOOLS	
	1995	1996
0 -11	5	3
11-20	6	6
21-30	3	2
31-40	5	5
41-50	3	4
51-60	0	2
61-74	2	2
TOTAL	24	24

High school administrators were asked to identify the number of grade levels involved in various types of multiple learning environments during the 1995-96 and 1996-97 school years. On average, 2.06 grade levels per school were providing out of class experiences in 1995-96; 2.28 were in 1996-97 (See Table 6). Schools, on average, did not support a single grade level for work site simulations, internships/apprenticeships, or youth-run enterprises in either school year. High schools averaged over two grade levels of employer visits, and over 1.5 grade levels of job shadowing and community service projects for each year of the program (1995-96 and 1996-97). Simulated work environments, school assignments in the workplace, youth-run enterprises, internships, work-linked learning, and work study co-ops did not average one grade per school and did not take place at all four grade levels in any of the schools.

On a 1 to 10 scale (lowest to highest), high school administrators were asked to estimate their commitment to staff development related to School to Work. Active/hands- on learning averaged 4.33 for 1995-96 and 5.11 for 1996-97, with a range of 0 to 9 and 0 to 8, respectively. Developing multiple learning environments, partnership strategies, world class standards, and integrating career development in 1995-96 and 1996-97 all had a mean score of less than 5. The exception was integrating career development, it did achieve a mean of 5.56 for the 1996-97 school year.

Most administrators did not report that a majority of the activities increased in occurrence from year one to year two (See Table 6). The most common activities that took place were out of class experiences, employer visits, and job shadowing. Each individual activity had an average range of occurrence of .5 to 2.44 grade levels. There was an average one grade level of support per School to Work activity for each high school.

Each high school administrator was asked to estimate his/her perception of the importance of different types of staff development. All of the staff development areas shown in Table 7 were emphasized in the federal legislation, as well as the regional partnership State-approved proposal, as necessary for all implementation. On a scale of 1 to 10 (lowest to highest), the highest average commitment to any single staff development component was the integration of career development, which had a mean level of commitment of 5.6. High school administrators demonstrated a low level of commitment to all staff development components related to School to Work. Partnership strategies, an activity used to develop relationships with local employers and organizations to get them involved in School to Work, had a level of commitment just above a 3.

There was a high level of agreement among high school administrators that School to Work should be a part of curriculum and instruction (See Table 8). High school administrators on average perceived School to Work as a more important activity than did superintendents. In contrast, high school administrators and superintendents agreed that technology should be a major component of School to Work. On a scale of 1 to 10 (lowest to highest), high school administrators also were asked to indicate their level of agreement with four statements about the 1997-98 school year. On average, School to Work as a part of high school curricular and instructional efforts was ranked 8.22. The actual range for responses from administrators to this statement was from 6 to 10. The statement that business and industry should be active in the development and integration of School to Work resulted in the mean response of 6.28, with a range of 1 to 10. The statement that community and parents should be active in the development and integration of School to Work had a response mean of 5.83, with a range of 1 to 10. Technology and electronic learning as a major component of School to Work received a response mean of 7.72, also with a range of 1 to 10.

Table 6
High School Administrators: Responses to Survey

	Number of Grade Levels			
	Min.	Max.	Sum	Mean
Out of class experiences				
1995-96	0	4	37	2.06
1996-97	0	4	41	2.28
Work-site used as classroom				
1995-96	0	2	6	.33
1996-97	0	2	10	.56
Internships or apprenticeships				
1995-96	0	2	12	.67
1996-97	0	2	12	.67
Co-ops of youth run enterprises				
1995-96	0	4	10	.56
1996-97	0	4	10	.56
Employer visits to classroom				
1995-96	0	4	39	2.17
1996-97	0	4	44	2.44
Job shadowing				
1995-96	0	4	28	1.56
1996-97	0	4	35	1.94
Simulated work environment				
1995-96	0	4	14	.78
1996-97	0	3	11	.61
School assignments in workplace				
1995-96	0	2	5	.28
1996-97	0	3	9	.50
Youth-run enterprises				
1995-96	0	2	6	.33
1996-97	0	2	6	.33
Community service projects				
1995-96	0	4	27	1.50

1996-97	0	4	30	1.67
Internships				
1995-96	0	2	11	.61
1996-97	0	2	10	.56
Work-linked learning				
1995-96	0	4	9	.50
1996-97	0	4	9	.50
Work-study co-ops				
1995-96	0	2	11	.61
1996-97	0	2	13	.72

Table 7
High School Administrators Responses: Commitment to STW Staff Development
(1 to 10 scale)

	Maximum	Mean
Active/hands-on learning*		
1995-96	9	4.3
1996-97	8	5.1
Development multiple learning environments		
1995-96	8	2.9
1996-97	8	3.4
Partnership strategies		
1995-96	9	3.2
1996-97	9	3.3
Develop world-class standards		
1995-96	10	3.6
1996-97	10	4.2
Integration of career development		
1995-96	9	4.4
1996-97		

*Response was 1 for all items

Table 8
High School Administrators' Response: Actions for 1997-98

	Low	High	Mean
STW should be part of curriculum and instruction	6	10	8.2
Business and industry should be active in integration	1	10	6.3
Community and parents should be involved in integration	1	10	5.8
Technology should be a major component of STW	1	10	7.7

* level of agreement 1=none 10=completely

BEST COPY AVAILABLE

High School Administrators: Analysis of Responses as a STW Implementation Profile

As shown in Table 9, there is a strong positive association between the high school administrators' School to Work implementation profile and the amount of job shadowing and simulated work environment activities. The more job shadowing that was taking place in the high school, the higher the overall level of implementation of School to Work. There is no association between the high school administrator's and the superintendent's estimation of the level of implementation of School to Work.

High school administrators were asked to estimate their level of commitment to School to Work staff development activities as described in the Federal legislation and the State-approved regional partnership proposal for the implementation of School to Work policy. Based on high school administrators' responses, there is a high association between the level of implementation of School to Work and the administrators' commitment to active/hands-on learning, development of multiple learning environments, utilization of partnership strategies, development of world class standards, and integration of career development staff development activities (See Table 10). Table 11 shows the positive high correlation between the High School Administrators' School to Work Implementation Profile and the percentage of youth-at-risk, the number of years the superintendent has been employed in that title in the current district, and the total number of years the superintendent has held that title.

BEST COPY AVAILABLE

Table 9
High School Administrators: STW Profile & Specific Responses

High School Administrators	Supers Rank	Job Shadow	Simul. Environ.
Admin Profile Of STW '95	.320	.546*	.537*
Admin Profile Of STW '96	.451	.735*	.575*

	School assign in workplace	Youth Enterp.	Comm. Serv.	Intern-ships
Admin Profile Of STW '95	.305	.086	.570*	.331
Admin Profile Of STW '96	.221	.029	.439	.335

*p<.05 **p<.01

Table 10
High School Administrators: Commitment to STW Staff Development

	Active/ hands-on learning	Develop multiple learn environ.	Partn-ership Strat.	Develop World Class Standards	Integrate Career Develop.
Admin STW Profile in '95	.828**	.907**	.765**	.839**	.758**
Admin STW Profile in '96	.774**	.931**	.698**	.819**	.461

* p< .05 ** p< .01

Table 11
High School Administrators Profile of STW Implementation

	% YAR 1995	Superintendent: Yrs in district	Superintendent: Total # of yrs
HS Admin STW Profile 1996	.442*	.623*	.792**

* p< .05 ** p< .01

Regression analysis, using the high school administrators survey data, was used to determine the level of variance in high school administrators' perceptions of School to Work implementation explained by the performance, environment, and technology implementation profiles. The results are different from the analysis using the superintendents' responses. The level of variance in the High School Administrators' School to Work Implementation Profile is explained significantly ($p < .10$) by the high school's environment, performance, and level of implementation of technology. Roughly 50% of the variance can be explained with less than ten percent chance for error *(See Table 12).*

```
                         Table 12

                    Regression Analysis

Predictors: Environment Profile, Technology Profile,
            Performance Profile

Depend. var.: H.S. Administrators Profile of STW Implementation '95

 R=.707  Rsquare=.499           F=2.99          sig=.088

Depend. var.: H.S. Administrators Profile of STW Implementation '96

 R=.718  Rsquare=.515           F=3.184         sig=.077
```

High School Administrators: Summary

According to these high school administrators, the level of participation in School to Work increased from the 1995-96 to the 1996-97 school year. Each of the required School to Work activities was taking place in at least one grade level in every school in the region. In terms of staff development, integration of career development had the highest level of commitment. The development of partnership strategies, however, was among the lowest concerns. High school administrators saw a need for involvement in workforce development activities and the staff development for developing those activities in general. But they did not follow through with the training of staff or program implementation across all grades.

High school administrators extended strong support for the continuation of School to Work into the 1997-98 school year. In addition, it was important business and industry and community members remain involved in the implementation process. They also wanted technology to be a major component of the activities supported.

The higher the percentage of youth-at-risk in the school district, the higher the implementation of School to Work as described by the high school administrator. The greater the length of time the superintendent was employed by the district in his/her current title, the greater the level of School to Work implementation. The greater the total number of years the superintendent has held that title, the greater the level of School to Work implementation that took place in the high school.

Student performance, the local environment, and the level of technology implementation reached in the high school helped to explain the variation in the level of School to Work implementation perceived by the high school administrators. Specifically, there was a strong and significant relationship between the local conditions under which the school was operating while implementing the policy and the success of actual implementation.

School Selection and Follow- up Interview Process

To understand more completely what affected the school district participation in School to Work, follow-up interview methods were used. The intent of the interview was to gather data to determine why schools implemented the School to Work policy at varying degrees although they were part of the same consortium. High schools in the region were selected based on their performance profile and School to Work implementation profile, determining the school's placement on a grid (See Figure 1). The performance and School to Work implementation profiles were chosen as selection criteria because of the availability of data.

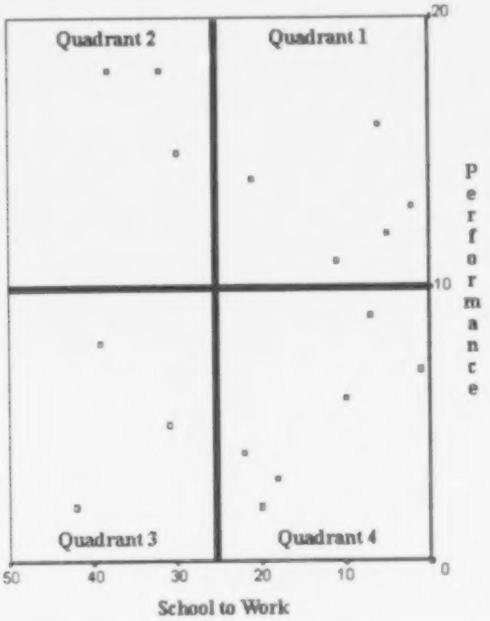


Figure 1. Placement of School Districts on the Grid

Summary of the Interview Process

The interviews with stakeholders provided information that could not have been collected through surveys. Interviewees shared what had happened during the decision making process when considering involvement in School to Work. All agreed the initiative was worthwhile, but there were common issues keeping schools from becoming involved: lack of resources to become involved, lack of resources for continuation, and competition with other efforts. Participating schools saw the School to Work initiatives as a means for schools to prepare and engage students in career preparation decisions. They also recognized the desirability of creating connections between the community and teachers, thus exposing them to new ideas.

Summary of Three Stages

Four profiles were created to determine to what extent characteristics of the local organization explained the variance in implementation of Federal policy. The performance, environment, and technology implementation profile of a local organization significantly predict the variation in School to Work implementation, when the high school administrator described the implementation. According to the survey results of superintendents, high school administrators, high school teachers, and employers, School to Work implementation levels increased from 1995-96 to the 1996-97 school year. During the same period, commitment to School to Work activities also increased. The majority of stakeholders surveyed felt very strongly that School to Work should continue during the 1997-98 school year.

Conclusions

Summary of Findings

This study compared the original intent of a specific educational reform policy to its local adaptation. Specifically, it focused on the implementation of the Federal School to Work Opportunities Act of 1994 in the 44 public high schools participating in the New York State Mohawk Valley Workforce Preparation Consortium during the 1995-96 school year. To compare the original intent of the School to Work policy to the local program adaptation and explain their differences two questions were answered. Participation in a consortium arrangement designed to facilitate the implementation of State or Federal policy resulted in active, if minimal, implementation efforts by individual consortium members. Stakeholders from the high schools studied described their schools as having been active in the implementation of School to Work during the 1995-96 school year. However, the level of implementation throughout the region was limited and minimal in some high schools.

The organizational and community context descriptions of those high schools in which School to Work was perceived as having both specific and important policy effects differed significantly from those where the policy was perceived as having little or unimportant effects. The greater the perceived effect of School to Work on operations, the higher the level of School to Work implementation in both 1995-96 and 1996-97. The greater the perceived importance of School to Work, the higher the level of implementation. However, in no high school had the program been implemented in all four high school grades. A high school setting where School to Work activities were perceived by local stakeholders as having great specific and important effects differs from a high school setting where School to Work activities were perceived as having little or no effect. The differences in organizational demographics and community characteristics that contributed to the high school administrator's description of School to Work implementation are statistically significant. Profiles of high school student performance, high school operating environment, and the level of technology implemented predicted the level of School to Work implementation described by the high school administrator. However, none of the individual profiles explained the variance in School to Work implementation by themselves. Therefore, the extent to which there were differences in perceptions of the enriched or impoverished status of technology did not contribute to perceptions of School to Work. Similarly, there were no significant differences in perceptions of curriculum scope and sequencing or in instructional strategies seen as being influenced by School to Work.

Based on the findings of this study, implementation was assessed by describing the reported activities of the local integration and comparing them to the original intent of the policy. The difference between original intent and reported local adaptation were interpreted through the backward mapping analysis as discrepancies in fulfilling the purpose of the particular policy. This study argued a more inclusive approach.

Backward mapping includes an explanation of local conditions and expands policy analysis beyond the determination of success or failure of a policy based on the measurement of the objectives. This bottom-up approach explained the conditions under which the policy was implemented. The process was more than simply measuring success by meeting policy goals; it explained the relationship between characteristics of the organization and local level policy implementation. Thus, there was consideration for the context in which implementation took place. In terms of context, backward mapping delineated both the conditions prior to and during implementation and their relationship to the variance in the level of implementation.

Backward Mapping as a Method of Policy Analysis

Backward mapping was effective in accomplishing three important objectives: analyzing policies that have been created to have an impact at the local level, providing contributions to the knowledge of the policy implementation process, and enabling a determination and explanation of the success of the studied policy. The process takes into account the difference in localities. Therefore, the analyst takes into consideration that the many differences between the localities--including leadership, environment, student performance, involvement in other innovations, and stakeholder perceptions--contribute to the differences in policy implementation among these localities. Consequently, when the implementation process is to take place over multiple years, or if it is slated for replication in other localities, the analyzed policy can be adjusted or improved based upon the findings. Forward mapping, as a comparison, explains to what extent the implementor has met established benchmarks. Variation in implementation is only considered more or less successful compared to a set goal, which is frequently complete integration. Backward mapping takes the analysis an additional step and explains the extent to which the locality's perceptions influenced variations of implementation. By addressing perceptions and the conditions of the environment, backward mapping removed the focus from the policy objectives in determining implementation success. Rather, success is expressed in terms of the effect of the local implementing organization. Backward mapping was useful for determining and explaining the relationship between the locality prior to policy implementation and how policy, as intended to be implemented by the federal government, was adapted to those conditions. During the process the direct analysis of the policy was shifted to focus on the high school. The organization's impact on the policy guiding the implementation was explained through an analysis of its characteristics. The result was the delineation of predictors associated with the successful implementation of the policy.

However, the comparison of intent and actual implementation is only one step in the backward mapping process. Backward mapping is also useful for analyzing why the implementation did or did not take place. Answering this question requires a process which enables the researcher to examine the conditions surrounding the policy. Because schools do not operate in a static environment, the policies being implemented by them are subject to forces which can cause the school itself to change. Simply measuring the ability to meet objectives is often worthwhile and appropriate. But for explaining varying degrees of implementation and how mediating variables may be dealt with by future policies, the information compiled in the backward mapping process is beneficial.

Student performance, environment, and technological characteristics of the high school explained the variation between Federal intent and local adaptation. Stakeholders were asked to provide their perceptions of the implementation process. Concurrently, the data collected about the organization explained what had been taking place at the local level, irrelevant of the policy implementation. For example, the organization may have been involved with other important and specific challenges that precluded an implementation at a pace similar to other organizations. Likewise, similar groupings of organizations may have differentiated levels of implementation due to a common factor, such as the participation in a particular event. Backward mapping was an effective method for delineating that implementation variance was explained by characteristics of the local organization.

Backward mapping ascertained that organizational leadership, past performance, stakeholder perceptions, external environment, and involvement in other efforts were all associated with policy implementation success. Both forward and backward mapping explained the level of implementation attained. But backward mapping also explained why the levels of implementation were reached. Backward mapping, therefore, was an effective tool in explaining the relationship between the original policy, the implementing locality, and the actual implementation.

Backward Mapping and Local Level Characteristics

Backward mapping elicited information which supports an understanding that successful implementation is not based solely on well-written policy. Characteristics of the high school which impacted the implementation process were identified and measured. Their association with variation of school implementation from established goals was explained in the final analysis of the process.

Perceptions and Participation

It was found that the district and building level leadership had an impact on the level of policy implementation. Each leader, as a stakeholder, had a different perception of the policy's interpretation and level of importance. Participation in terms of the policy implementation meant signing as a consortium partner. But the variation of implementation among schools could not be explained simply by agreement to join a consortium. All of the schools identified themselves as partners in the submission of the grant to the State, but not all of them implemented the same level of School to Work initiatives.

Instead, interpretation of the policy influenced actual participation. Some school leaders interpreted School to Work as being career exploration. Those schools implemented basic activities such as tours, speakers, and job shadowing. But they were less likely to have implemented components of the policy that took a higher level of commitment of resources, such as work-based learning. At least two schools excluded themselves from initially participating because of their leaders' limited interpretation of the legislation. Not until there were definitive examples of the policy being implemented in other school buildings did these leaders "buy-into" the concept.

Decision Making

The original School to Work policy provided the flexibility to adapt implementation to meet local needs. It was at the discretion of the consortium to determine how the framework of support and implementation would be provided. The local school determined how the policy was actually implemented. This study found that the way in which a superintendent perceived multiple reform efforts, as well as the way the high school administrator perceived the School to Work policy, explained the discretion used to adapt the policy to local need. For some schools, it meant implementing School to Work components as described within the language of the legislation.

Other schools participated for different reasons. One school stated that it used School to Work support as a means for exposing its staff to new ideas and environments with the expectation that participation would effect change for the benefit of children. There is evidence that this happened, both in that school and others.

High school teachers involved in School to Work activities described changes in curriculum scope and sequence or in instructional strategies. Teachers described how classrooms and individual lessons had been affected positively because of their involvement in staff development programs funded by School to Work. In many cases, they became more aware of the resources (e.g., lessons being used by local teachers) around them. They discovered how the application of learning had a positive effect on the students. These teachers had previously believed that the application of knowledge did not have to take place in the classroom for students to learn and develop skills.

Teachers also discovered that there were many opportunities for students to test their knowledge in surroundings not previously considered learning environments, such as factories, farms, stores, and the community.

Competition for resources

While this study did not set out to analyze the issues of resources, there are some interesting findings are worth mentioning. The competition for scarce resources contributed to the variation of policy implementation levels between school districts. Poor schools described two areas of concern: the level of resources necessary to become involved and the resources needed to sustain the effort. Poor schools with a small local employment base that perceived School to Work merely as a career awareness strategy had low rates of participation. These educators did not see how they could become involved when there were so few community resources. Other poor schools were concerned about sustaining the effort after funding was no longer available. Either they did not feel the school had the fiscal resources to continue the initiative, or they perceived School to Work as one more initiative that would simply disappear.

Resource allocation also impacted the decisions of more affluent schools. Some of these schools and their communities put high demands on the students to perform academically. Their expectation of student preparation was for college and highly skilled careers. In some cases, this expectation resulted in School to Work implementation at higher levels. It was a means of exposing students to the demands of the workplace and the high level of skills required to be successful in the positions these students sought. For other schools with the same high expectation of their students, the perception of School to Work as applicable only for the "career-bound" was a detriment. Their perception was the student should focus solely on academic preparation leading to college.

Summary of Backward Mapping Policy Analysis

By using backward mapping in this study, attention was focused on the importance of the locality during the policy implementation process. Backward mapping established a broad framework within which an analysis of the actual implementation process was carried out. Within this process, the analyst is able to discover and explain what determines the level of success of the policy. By choosing any single Federal policy, one could determine the actual impact on the local school district. Using this process, it can also be determined the role of the locality in relation to the policy. The policy exists to create change at the local level, but it is the locality which ultimately decides how the policy's intent will be adapted to meet its needs. The key elements of backward mapping in policy analysis are as follows:

- It is effective in describing actual versus perceived implementation.

- It is effective in explaining actual versus perceived participation.
- The effect of the locality is considered to help to explain variation in implementation.
- Perceptions of importance and policy effect are measured.
- A more complete picture of policy implementation is presented.
- The thoroughness with which policy implementation information is provided is at a high level.

Backward Mapping, Forward Mapping, and School to Work

In the case of School to Work, a form of forward mapping analysis had already been completed by the Mohawk Regional Consortium, the New York State Education Department (NYSED), and the Federal government. Each level of governance had used previously set criterion established within the original Federal legislation, approved regional implementation plans, and state level implementation plans to determine the success of the policy. These studies found that the schools had the required representation in place and that they were engaged in the required activities. Thus, the state and the region appeared determined to have successfully implemented School to Work policy. Objectives of the policy had been met to varying degrees in each school, and continuation of the project was contingent on the objectives set for the following year of implementation. According to Mohawk Valley Consortium members, there had been little feedback given to the region from these studies upon which to base continuation on a commitment of resources to areas in need of improvement.

The data collected during this process that can be labeled as an example of forward mapping was for the most part unidimensional. There was snapshot of activities on a given day. But the discussion of the data collected assumed that all the activities were the result of School to Work participation and use of funds from the program. None of the data was in a format enabling the determination of association or any such relationship linking policy to actual implementation. Much of the data was compiled into anecdotal stories that spoke of program objectives being met in various schools.

There were 44 school districts with high schools that agreed to be partners in the Mohawk Valley Workforce Preparation Regional Consortium to implement the Federal School to Work Opportunities Act of 1994 (STWOA). Consistent with other non-academic studies involving this region, this study found that schools participated in varying degrees. Therefore, the agreement to participate in a consortium arrangement designed to facilitate the implementation of Federal policy did not result in identical implementation levels being achieved by individual consortium members.

By using backward mapping, it was found that the level of participation is best explained by the perceptions of the high school administrator. In addition, the perceptions of the high school administrator are predicted by the characteristics of the school. The high school administrators' perceptions of School to Work implementation vary based on the performance of the students, conditions in the community, and level of technology implemented in the school. The better the students are performing, the more resourceful the external environment. The more technology present in the school, the higher the level of policy implementation. Therefore, the level of school participation in the policy's initiatives varied according to perceptions and characteristics of the locality.

Local Characteristics

Characteristics of the local school districts in the Mohawk Valley Consortium were measured to explain the variation of high schools' involvement in School to Work policy initiatives. The school's characteristics and stakeholders' perceptions were measured to determine their impact on high school participation. By grouping characteristics into three broad areas called profiles (performance, environment, and technology implementation), the combination of profiles were found to explain the variations in success of School to Work implementation.

Performance Profile

The performance profile was used as a grouping that characterized student performance as a means for measuring the school's effectiveness. When the performance profile was combined with both the environment characteristic profile and the technology implementation levels profile, the variation of implementation between schools was explained to a high degree. Therefore, the combination of previous student performance, the conditions of the community, and the level of participation in other reform efforts combined predict the level of implementation of the School to Work policy.

The performance profile of the high school was found to have a few specific items that were significantly associated with the level of School to Work implementation. The greater the percentage of the total number of students receiving NYSED Regents Diplomas, the higher the level of School to Work implementation. This becomes an important statistic considering the perception of the teachers interviewed. Teachers attributed low implementation levels to the lack of flexibility of students working towards a Regents diploma. The teachers' cited specific problem areas, such as preparing students specifically for items on the Regents exam, and the lack of time for "extra- curricular" and exploratory activities that prevented Regents student participation. The teachers' statements are contradictory to the finding. Therefore, a question is raised as to what was happening in the classroom where teachers were asked to participate in School to Work but did not because of Regents requirement on students. Were the teachers' perceptions inaccurate? Were there a greater number of non-Regents students actively participating in School to Work and more School to Work activities in place in schools where higher percentages of students receive a Regents diploma? If so, could this be for the purpose of providing more opportunities in those schools for those non- Regents students if they were assumed to be workplace bound? These are questions to be addressed in further studies of the policy.

Schools within the Consortium with high percentages of Regents students and a high level of School to Work implementation had perceived School to Work to be something more than career exploration. In these schools, there was evidence of career awareness activities such as tours, speakers, and job shadowing all of which were typically already taking place, and which took less commitment by the teacher and the school to integrate. But there was also evidence of higher levels of integration taking place within the curriculum, instruction, and even the assessment. Schools graduating 85% or more students with Regents Diplomas had created interdisciplinary projects in areas such as Math, Science and Technology. They also changed methodologies in the classroom, incorporating applied learning techniques and utilizing the community and the workplace as learning environments. Schools had even begun to assess student knowledge through hands-on application and problem solving using actual workplace situations. These schools did not perceive the Regents diploma as a barrier; they used School to Work initiatives as a means for creating new ways to meet its rigors. According to this study's findings, these occurrences can be attributed to two factors: the leadership provided by the high school administrator and the support of multiple reform efforts by the superintendent.

Environment Profile

The environment profile was a grouping which characterized the conditions of the environment in which the local high school operated. Schools located in communities of high and low unemployment and wealth were able to implement School to Work policy initiatives. Measures of association were applied to determine the extent to which the level of School to Work implementation was explained by specific community characteristics. It was found that the types of business in the community and the percentage of youth at risk were highly correlated to levels of implementation. The higher the percentage of people in the community employed by agriculturally-based businesses, the higher the level of implementation of School to Work. Even though agriculture was the lowest percentage of the total businesses, no such significant correlation could be found for manufacturing, retail, or service-based businesses. The analysis also determined that the higher the percentage of youth who are "at-risk", the higher the level of School to Work implementation. Several factors were discovered during the quantitative follow-up interview process that could explain this fact. Prior to the implementation of the STWOA of 1994, several State and Federally funded programs were addressing the needs of the youth-at-risk population through School to Work-type activities. The Governor's School And Business Alliance (SABA), Private Industry Councils (PICs), and Workforce Preparation Pilot projects were all based heavily on the needs of high school children from low wealth families. Twenty-seven of the 44 Mohawk Valley Workforce Preparation Consortium school districts were located in a BOCES service area that also had a SABA. Fifteen of the 44 school districts were involved in the original New York State Workforce Preparation Pilots funded through the Governor's office. Twenty-four of 44 school districts were located in a BOCES that was involved in the second round of Workforce Preparation grants funded by the Governor's office and managed by NYSED. All of the schools were supported by a regional PIC. Therefore, all of the schools had been participating in a combination of School to Work-type initiatives based on their at-risk youth population prior to the implementation of STWOA of 1994.

Technology Profile

The third local characteristic groupings to be measured were those related to participation in local reform efforts. There were actually two measurements of commitment to other reform efforts. One was the level of implementation of educational technology within the school, which was determined from an existing data set. The second was stakeholder perceptions of the importance of involvement in other reform efforts that was collected from this study's survey. The level of technology implemented in a school was not significantly associated with the level of School to Work implementation. However, the more important the involvement in multiple reform efforts, the higher the level of technology integration into a school. There was found to be a strong association between technology implementation level and the level of importance of participation on multiple reform efforts as indicated by the school superintendent. In addition, there was a strong and significant association between the level of School to Work implementation as perceived by teachers and the level of importance of participation in multiple reform efforts as indicated by superintendents. The more important a superintendent perceived participation in multiple reform efforts, the more School to Work activities and technology were implemented by the teachers. Therefore, it was not the actual participation in another reform effort that explained future participation in other reform efforts. It was the perception of the superintendent of the importance of involvement in other reform efforts that was associated with implementation of other Federal policies.

Stakeholders' perceptions of involvement in multiple reform efforts were measured as a local characteristic that was associated with the level of School to Work implementation. The perceptions of the superintendent were not indicative of the level of School to Work implementation in the high school. The analysis of local level data did find that there was a significant and very strong relationship between the number of years the superintendent has been employed in the district under his/her current title, the total number of years he/she has held the title, and the level of School to Work implementation as reported by the high school administrator. Therefore, the longer the person held the title of superintendent, the more likely School to Work was reported by the high school administrator to have been implemented.

In comparing the perceptions of educators and employers involved in School to Work, it was found that neither superintendent's nor employers' perception of an effect on organizational operations was associated with involvement. This can be attributed superintendents perception of School to Work as means for creating change in the classroom and providing opportunities for students, rather than as a means for impacting the operations of the organization. But part of the impetus for the region to become involved in School to Work was to work with employers to improve the future workforce. However, the results showed teachers knew very little of the partnerships that existed with business. There was ample documentation in the region of partnerships including directories of partners, teacher/student documented activities with business, and a computerized database of available employer partners and the types of activities they would support. But the teachers and administrators did not identify them as existing in their building.

It was assumed, based on the available literature, that businesses were involved, at least partly, because they were concerned for their future employees and how well they were being prepared. However, employers participated in the activities for reasons other than organizational impact. During the follow-up interview process employer-partners indicated there had not been enough time invested in the implementation of School to Work to realize impact on their own business. But it was expected by employer-partners interviewed that employers would be able to see an impact in the long term.

Continuation of Implementation

Most schools indicated that it would be important to participate in School to Work activities during the 1997-98 school year. The level of previous implementation was not associated with the level of stated importance of participating the following year. A majority of the schools indicated a high level of agreement that the integration of School to Work into curriculum and instruction with the involvement of community members and employers was very important.

The business and industry members of the community described School to Work involvement during 1995-1996 as being important. All of the respondents strongly agreed that School to Work should continue, but at the same time they did not perceive the policy as having any impact on the operations of the business. A few employer partners did expand on this point. Their view was that School to Work is a means for helping students learn, which was their rationale for participating. There would need to be a longer-term interaction before they could assume that the policy would have an impact on the operations of the business. In the long-run they did assume that students would be better prepared for work with such programs. For that level of impact to take place, they asserted that the School to Work implementation would have to be an integral part of all schools.

Local Characteristics Summary

By using backward mapping this study found that the perceptions of stakeholders in district and building level educational leadership positions were an important factor in determining the level of implementation of School to Work policy in 9-12 high schools located in the Mohawk Valley Region. The high school administrators' reported level of implementation varied according to local characteristics and conditions. When superintendents found involvement in multiple reform efforts to be important, teachers reported higher levels of School to Work implementation. Actual implementation was best predicted by the high school administrator. Yet, only one BOCES in the region documented providing technical assistance and training specifically to high school administrators. Had the significance of the high school administrator been known the region and its leadership may have placed greater emphasis on staff development for high school administrators. It may have marked resources dedicated to improving the leadership and understanding of School to Work and prepared them to provide more effective support during the implementation of this policy. This should not be considered a weakness on the part of the regional leadership, but a lack of information available to policymakers who required superintendents to sign-off and teachers to implement School to Work. Backward mapping has provided the means for collecting this information that, if used, could impact the School to Work implementation policy for the last two of the five-year policy implementation phase in New York State and subsequent years in other states.

Using Backward Mapping Process for Policy Analysis

Policy analysis through a backward mapping lens has proven to be an effective process for determining the predictors and identifying items that explain the success of implementation. Previous analyses of School to Work implementation in the Mohawk Valley Region followed a forward mapping and top-down process. These attempts to analyze implementation of the Federal policy determined the level to which the region met goals and objectives and how much School to Work was implemented. It did not explain why the implementation took place, or who or what contributed to the level of implementation. The results of the forward mapping process did not explain the variation of implementation levels between schools. The forward mapping process focused on the measurement of the volume of activities and the documentation of the involvement of stakeholder groups. It is the backward mapping approach which addressed how their involvement impacted decision making and final implementation in the form of local adaptation.

The backward map explained first, that schools participated at varying levels and secondly, why schools participated at varying levels. To reach these findings, the process required an initial analysis to determine that there was in fact a difference between the schools and the environments in which they operated. A determination was then needed that each school agreed to and followed the same guidelines, adapting the initiatives to meet it's students needs. Each individual school then participated at a level which best suited it's stakeholders. The stakeholders' decisions and perceptions impacting implementation are attributed to a combination of pre-existing characteristics of the school's internal and external environment. Pre-existing conditions in and around the school and the perceptions of stakeholders were measured by the backward mapping process, and resulted in an awareness that such actions are valuable to policy analysis. If the policy analysis process had stopped at measuring goals and objectives, there would have been no determination of the reasons some schools contributed to meeting those objectives. It would not have explained how the attainment of some objectives was actually due to schools participating in other initiatives.

It is the backward mapping of the policy that enables as analysis recognizing that the locality may interpret the intent of the policy differently from that of the policymaker. Forward mapping assumes the failure to meet objectives is due to the locality's inability to implement. Backward mapping asks the local actor as implementor how they interpreted the policy. It assumes a variation in interpretation results in a variation in adaptation and implementation. If School to Work was locally defined as a career exploration program, then implementation would reflect activities to support student career exploration. In that case, school-based and work-based learning activities would be less likely to take place. Therefore, a school's lack of implementation is not related to its ability; it has to do with discretion based on perceptions. Choices were made because of forces acting upon the school that impact local actors' decision making.

The analysis was built from a local-level perspective to explain local actions to meet a Federal policy guideline. The survey instrument of the backward mapping process asked the district level, building level, classroom level, and employer partner to describe their perception of School to Work, what it meant to them, how important it was, and to what extent had it really taken place. To take the analysis a step further, the process used this data to identify special cases. These cases either met or went against expectations and were slated for a follow-up process. This second step looked for actual documentation and reviewed initial findings to gather a clearer understanding why certain perceptions were held and actions taken. The effort yielded a compilation of data that explained how it was the conditions within which the locality operated and made decisions that impacted implementation, not the clarity of goals or objectives. Therefore, future policy making should be based on more complete information on the locality in which the target population resides and for which policy is intended, not clearer policy goals.

Limitations of the Study

Limitations of the study can mostly be attributed to the use of the technology implementation profile. The technology implementation profile used the total number, rather than the ratio of computers to students or teachers in the school. This was problematic because a large school could have larger numbers of computers, yet have a lower student or teacher to computer ratio. A smaller school with a lower total number of computers but a higher student to computer ratio would have a comparably better level of technology integration. This may have skewed the rankings of this particular profile and may have impacted this profile's lack of explanation of variance in the School to Work implementation profile.

Another issue with the creation of the technology implementation profile was the time lag between the data collected for the Mann study and the data collected for this study. The time period was actually five months. Schools that were surveyed last had more time opportunity to install more computers. Therefore, data could have been skewed in favor of school districts that were surveyed later in the study.

Implications for Future Research and Policymaking

Backward mapping is not specific to any one type of policy. The process followed in this study could be replicated for any policy. In this study, backward mapping was successfully applied to a Federal policy implemented at the local level. The actual items specifically measured related to School to Work policy could be replicated in other areas of New York State. There are over 50 other partnerships throughout the state for which this process could be useful in determining factors of local success and further testing the application of the backward mapping process. There were seven other states that began implementation during the same time period, and several more that have begun implementation in the last year. All of these local partnerships could be analyzed using this approach. The specifics of replicating this particular study could be challenging. Student performance items from other states would have to be substituted with that particular state's student performance indicators. Therefore, the use of backward mapping is universal in its application to other policies. The exception would be in cases where local data was not available or the analysis does not warrant the consideration of local action. In cases where there is no means or cause for the use of discretion by the local actor, backward mapping may not be the most appropriate method.

The creation of future policy could be enhanced through the use of results from a backward mapping policy analysis. The findings contribute to an improved understanding of how the implementing organization and the environment in which it operates react to the policy. This seems to be most useful in cases of multi-year implementation policies. These are situations where the policy is implemented, the policy can be analyzed after the first year, and adjusted to improve the effectiveness of the policy in future years. Short-term policy implementations would most likely not yield information useful for the improvement of policy implementation in that specific case. However, it would be useful for replication of the policy implementation in other states or localities and provide useful information for the improve of other future policies. It also would be effective in comparing implementation levels during funding years and continuation levels after funding has run out. Essentially, backward mapping provides for a more global understanding of how policy implementation works and what impacts its success.

References

Bach, B.W. (1989). The Effect of Multiplex Relationships Upon Innovation Adoption: A Reconsideration of Rogers' Model. *Communication Monographs,56,2.*133-150.

Benz, M.R., et al. (1997). School to Work Components That Predict Postschool Success for Students with and Without Disabilities. *Exceptional Children,63,2.* 151-165.

Berman, P. and McLaughlin, M.W. (1973). *Implementing Innovations: Revisions for an Agenda for a Study of Change Agent Programs in Education. Study of Change Agent Programs: A Working Note.* RAND Corporation. Santa Monica.

Coleman, J.S., and Dobeare, K.M. (Ed). (1975). Problems of Conceptualization and Measurement in Studying Policy Impacts. *Public Policy Evaluation.* Sage. California.

Council of Chief State School Officers. (1996). 1996 CCSSO Issues and Positions: Federal Education Policy and Funding. (On-line). http://www.ccsso.org/96legpap.htm.

Davis, O.L., Jr. (1994). Don't Adopt; Adapt; A Reminder for Every Year. *Journal Of Curriculum and Supervision, 9,4.* 403-405.

Decker, P.T. (1997). *Education and the Economy: An Indicators Report.* Mathematica Policy Research, April.

Dunn, W.N. (1994). *Public Policy Analysis: An Introduction, 2nd Ed.* Prentice Hall. New Jersey.

Education Reform and School to Work Transition In Washington State. (1997). (On-line). http://www.wa.gov/wtb/edreform.html

Elmore, R.F. (1978). Organizational Models of Social Program Implementation. *Public Policy, 26,2.*185- 228.

Elmore, R. F. (1979). Backward Mapping: Implementation Research and Policy Decisions. *Political Science Quarterly, 94.* 601-616.

Elmore, R. F. (1987). Instruments and Strategy In Public Policy. *Policy Studies Review, 7, 1.* 174-186.

Elmore, R.F., and McLaughlin, M.W. (1988). *Steady Work: Policy, Practice, and the Reform of American Education.* The RAND Publication Series, February. RAND Corporation. Santa Monica.

Elmore, R. F, and Williams, Walter, (Ed.). (1982). *Studying Implementation.* New Jersey: Chatham House.

Ely, D.P. (1990). Conditions that Facilitate the Implementation of Educational Technology Innovations. *Journal of Research and Computing in Education, 23,2.* 298-305.

George, C.A. (1993). Strategic Planning: Key Environmental Indicators for Pennsylvania. *Pennsylvania Educational Policy Studies, 17.*

Giesecke, J. (1993). Recognizing Multiple Decision- Making Models: A Guide for Mangers. *College and Research Libraries, 54,2.* 103-114.

Giesecke, J. (1994). Recognizing and Managing Multiple Organizational Approaches. *The Dynamic Library Organizations in a Changing Environment.* 29-46.

Grummon, P. T.H. (1994). *Evaluating Systemic Change in School to Work Initiatives.* Paper presented at Annual Meeting of the American Educational research Association, New Orleans, April.

Hall, G. and Galluzzo, G. (1991). Changing Policy Into Practice: School-based Decisionmaking. *Policy Issues, November.* State Policy Program: Appalachia Education Laboratory.

Hess, M.A. (1997). School to Work: Linking Learning to Livelihoods. *Curriculum Update. Fall 1997.*

Ingram, H., and Schneider, A. (1990). Improving Implementation Through Smarter Statutes. *Journal of Public Policy, 10,1,* 69-70.

Jennings, J. H. (Ed.). (1995). *National Issues In Education: GOALS 2000 and School to Work.* Washington, D.C.: Phi Delta Kappa International and The Institute for Educational Leadership.

Johnston, B.J. (1993). The Transformation of Work and Educational Reform Policy. *American Educational Research Journal, 30,1.* 39-65.

Legislative Analyst's Office. (1995). Implementing New Federal Education Legislation. (On-line). http://www.lao.ca.gov/rp020195.html.

Lindblom, C. E. and Woodhouse, E. J. (1993). *The Policy Making Process.* New Jersey. Simon and Schuster.

Mazmanian, D.A., and Sabatier, P.A. (1981). *Effective Policy*

Implementation. Lexington Books, Massachusetts.

McDonnell, L.M., and Elmore, R.F. (1987). Getting the Job Done: Alternative Policy Instruments. *Education Evaluation and Policy Analysis, 9,2.* 133-152.

McLaughlin, M.W. (1987). Learning From Experience. *Education Evaluation and Policy Analysis. 9,2,* 171- 178.

Mohawk Valley Workforce Preparation System Consortium. (1995). *Application for a Subgrant to Local Partnership under the New York State Education Department School to Work Opportunities Implementation Grant.*

Murnane, R.J. (1987). Improving Education Indicators and Economic Indicators: the Same Problems?. *Educational Evaluation and Policy Analysis, 9,2.* 101-116.

National Center on Education and the Economy. (1990). *America's Choice: High Skills or Low Wages.* National Center on Education and the Economy Commission on the Skills of the American Workforce. Rochester.

New York Association of Training and Employment Professionals. (1996). *Policy Frameworks for New York's Workforce Development System.*

New York State Education Department. (1995). *Planning and Implementation of a School to Work Opportunities System: Guidelines and Application Materials.* Office of Workforce Preparation and Continuing Education.

New York State Education Department. (1997). New York State School to Work: The Evaluation Program. (On- line). http://www.nysed.gov/workforce/stweval.html

New York The State of Learning. (1995). *Building a School to Work Opportunities System in New York State: New York's Application for an Implementation Grant Under the School to Work Opportunities Act.*

Oakes, J. (1986). Educational Indicators: A guide for Policymakers. *Center for Policy Research in Education, Occasional Paper Series, Oct, 1986.*

Oakes, J. (1989). What Educational Indicators? The Case for Assessing the School Context. *Educational Evaluation and Policy Analysis, 11,2.* 181-199.

Ormrod, R.K. (1990). Adaptation and Cultural Diffusion. *Journal of Geography, November/December.* p. 258- 262.

Orr, M.T. (1995). Evaluating School to Work Transition. Education Reform and School to Work Transition Series. *Academy for Education Development, National Institute for Work and Learning.* Washington, D.C.

Pressman, J. L., and Wildavsky, A. *Implementation.* Los Angeles, CA: University of California Press.

Report to Congress. (1997). *Background on the School to Work Opportunities Act.* (On-line). http://www.stw.ed.gov/congress/CG_TOC.htm.

Rogers, E.M. and Valente, T.M. (1995). The Origins and Development of the Diffusion of Innovations Paradigm as an Example of Scientific Growth. *Science Communication, 16,3.* 242-273.

Rogers, E. M., and Van de Ven, A. H. (1988). Innovations and Organizations: Critical Perspectives. *Communication Research 15,5,* p.632-651.

Sabatier, P. A. Top Down and Bottom-Up Approaches to Implementation Research: A Critical Analysis and Suggested Synthesis. *Journal of Public Policy. 6, 1,* pp.30- 36.

Scriven. M. (1993). Hard-Won Lessons In Program Evaluation. *New Directions for Program Evaluation, American Evaluation Association, 58, Summer.* Jossey- Bass. San Francisco.

Simon, H. A. (1993). Decision Making: Rational, Nonrational, and Irrational. *Education Administration Quarterly. 29,3,* 392-411.

School to Work in new Hampshire. (1997). About School to Work in New Hampshire. (On-line). http://webster.state.nh.us/doe/stwabout.htm#activ

School To Work Opportunities Act of 1994, Public Law 103-239, 108 Stat 568, May 4, 1994, H.R. 2884, Section 3, Washington D.C..

State of New York. (1995). In Assembly: 1995-96 Regular Sessions. June 9, 1995. 8106.

U.S. Department of Education. (1983). *A Nation at Risk: A Report to the Nation from the Secretary of Education.* The National Commission on Excellence in Education. Washington, D.C.

U.S. Department of Education. (1996). Progress of Education in the United States of America- 990 through 1994: Reform at the Local Level. (On-line). http://www.ed.gov/pubs/Prog95/pt3local.html

U.S. Department of Education. (1996). School to Work Initiatives

October 1996: Executive Summary. (On-line). http://www.ed.gov/pubs/SER/SchoolWork/study3.html.

U.S. Department of Education. (1996). School to Work Initiatives October 1996: Study Aims and Questions. (On- line). http://www.ed.gov/pubs/SER/SchoolWork/study3.html.

U.S. General Accounting Office. (1993). Systemwide Education Reform: Federal Leadership Could Facilitate District-level Efforts. *Report to Congressional Requestors. GAO/HRD-93-97.*

Weiss, C.H. (1977). *Evaluating Action Programs: Readings in Social Action and Education.* Allyn and Bacon. Boston.

Weiss, C.H., and Bucuvalas, M.J. (1980). *Social Science Research and Decision Making.* Columbia University Press. New York.

Wiles, David K. (1996). *Networking High Performance In New York's Secondary Education: The Regents Curriculum Story.* University Press of America. New York.

Windham, D.M. (1988). Effectiveness Indicators in the Economic Analysis of Educational Activities. *International Journal of Education Research, 12,6.* 575-665.

Windham, D.M. (1990). *Improving the Efficiency of Educational Systems: Indicators of Educational Effectiveness and Efficiency.* United States Agency for International Development, Office of Education.

Note

1. The Hamilton-Fulton-Montgomery BOCES was not a part of the original Dale Mann study because it is not part of the Mohawk Regional Information Center. The same questions and procedures were followed to replicate the technology survey study portion that provided pertinent information as to the level of technology implemented for this study.

About the Author

Arthur M. Recesso
Assistant Professor
Department of Secondary Education, Curriculum, and Instructional Technology
1500 N Patterson Street
Valdosta State University
Valdosta, Ga 31698

Phone: 912-259-5099 Fax: 912-333-7167
Email: amrecess@valdosta.edu

Dr. Recesso serves as Assistant Professor of Instructional Technology at Valdosta State University. He received his Ed.D. from the University at Albany, State University of New York. His primary research interests are in the fields of education policy analysis and instructional/educational technology. Dr. Recesso's recent activities may be found at http://education.valdosta.edu/dept/sedit/ and http://innovativeknowledge.com

---

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa- DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 190 times since April 10, 1999

Education Policy Analysis Archives

Volume 7 Number 12 April 10, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Hiring Policy in United States Spanish Departments: Considerations of Social Class, National Origin, and Ethnicity

Jerry Hoeg
Eric Cohen
Christine Fullen
Penn State University

Abstract

The present study focuses on two inter-related factors specific to United States college and university Spanish Departments: the unique demographic profile of the entry-level faculty in terms of gender, ethnicity, national origin, and social class; the relation between these factors and hiring practices, especially regarding field of specialization, pay scale, and tenure-track opportunities. We believe these issues are important in that they underscore the value of considering questions of social class as well as ethnicity and gender when analyzing academic job segmentation.

The present study focused on the initial job searches of recent recipients of the Ph.D. in Spanish from US institutions. It was conducted by mailing structured questionnaires to 150 of those listed in the 1997 "Dissertations 1996" (Eutis, 1997) section of *Hispania*, the journal of the American Association of Teachers of Spanish and Portuguese. This is an annual listing of completed dissertations in Spanish. In addition to follow-up telephone calls to degree-granting institutions, the 1997 *Modern Language Association Member Directory* and the "Job Tracks: Who Got Hired Where, 1997- 1998" listings in *Lingua Franca* were consulted, and surveys mailed where appropriate.

Results

A total of 99 people (out of 170, 58%) returned completed questionnaires. (Note 1) A comparison of our sample with the total universe of doctorates granted in Spanish (National Opinion Research Center, 1997) in 1996 (n=196; this is the year most of our sample received their degrees) across key descriptive variables displays relative correspondence, with almost all categories varying by three percent or less (Table 1). The Modern Language Association's PhD survey for the years 1993-1994 (Huber, 1996), the most recent data available, further supports the representation of the sample in terms of the proportion who received tenured or non-tenured positions (Table 1-A). Based on the close correspondence between the two populations, any concerns regarding selectivity or sampling bias are unfounded.

Table 1
Comparison of Survey Population and 1996 Universe

	Survey		1996 (Note 7)	
	n	%	n	%
Total	99	100	196	100
Male	42	42	82	42
Female	58	58	114	58
US Citizen	59	59	116	59
Perm. Visa	19	19	42	21
Temp Visa	15	15	34	17
Hispanic	48	48	98	50
Mex/Am	6	6	10	5
Puerto Rican	4	4	15	8
Other Hisp.	38	38	73	37
Male	21	21	38	19
Female	27	27	60	31
US citizen	14	14	28	14
Perm visa	19	19	31	16
Temp visa	14	14	24	12
Non-Hispanic	51	51	98	50
Male	21	21	44	22
Female	30	30	54	28
US citizen	45	45	73	37
Perm visa	1	1	11	6
Temp visa	4	4	10	5

Table 1-A
Comparison of Post-Secondary Education Placement Among Survey Population and MLA Placement Survey

	Survey		**MLA** (Note 8)	
	n	%	n	%
Total	99	100	222	100
Post-Secondary	78	78	182	82
Hispanic	36	46	87	48
Non-Hispanic	42	54	95	52
Tenure	50	64	124	66
Male Tenure	22	28	50	27
Female Tenure	28	36	73	39
Non-Tenure	26	26	41	22
Male Non-tenure	8	10	21	11
Female Non-Tenure	20	25	41	22

Several factors that do not appear in the Tables, but that might influence a job search, should be mentioned. First, only six of the Hispanics were native-born (eight were naturalized) and, if we correct for Puerto Ricans, who are native speakers and non-immigrants, only two of the Hispanics are potentially second- generation US citizens, hence non-native speakers. (Note 2) Also, the survey cohort ranged in age from 28-60 years of age. The mean age of the respondents was 36 (37 years for Hispanics, 36 years for non-Hispanics). Additionally, because our sample were on their initial job search, years of academic experience was more or less a constant in the present analysis. Almost 90 percent (88%) had one year or less with a doctorate. There appeared to be no pattern among the remaining 12 percent.

Lastly, publication of refereed journal articles is one significant indicator of merit or productivity which could lead to a more successful academic outcome. Approximately 65 percent of the sample had no peer-reviewed journal publications; 85 percent had one or less. Only 15 percent had 2 or more articles published, a number which could have a significant bearing on acquiring a tenure-track position. No pattern emerged between Hispanicity or gender and number of publications. Thus this variable is treated as a constant in the present analysis.

A oneway ANOVA was computed for Hispanic versus non-Hispanic differences and gender subgroups (see Table 2). Based on the ANOVA results, entry-level Hispanics are paid more on average than non-Hispanics (t=2.69, df=67, p<.01). Differences between Hispanics and non-Hispanics based on gender were also significant (F=2.89, df=3, p <.05). The Hispanicity/gender effect was stronger than either the Hispanicity or gender influence taken separately. Mean differences in pay between Hispanic and non-Hispanic males were most pronounced, \$36,818 vs \$31,250 respectively (median differences = \$37,000 vs \$32,000).

Table 2
Tenure Status and Salaries (in $1000s) by Ethnicity and Gender

	Hispanic				Non-Hispanic					
	Male		Female		Male		Female		Total	
	n	% or Mean	n	% or Mean	n	% or Mean	n	% or Mean	n	% or Mean
Tenure*	11	84.6%	15	65.2%	11	64.7%	13	52.0%	50	64.1%
Aver. Pay**	11	$36.8	23	$34.9	14	$31.2	21	$32.5	69	$33.7
Lang.***a	0	—	6	24.0%	8	47.1	9	39.1%	23	29.5%
Lit.***	10	76.9%	10	40.0%	4	23.5%	6	26.1%	30	38.5%

* Sig. <.07
** Sig. <.05
*** Sig. .01
a. The abbreviations "Lang." and "Lit." refer to the area of teaching specialization.

For those who were offered full-time academic jobs (n=78), a comparison was made between those offered positions where the primary focus was teaching literature versus those whose primary responsibility was teaching language. Differences between Hispanics and non-Hispanics showed a weak statistical significance. When the sample was broken down by gender subgroups, differences were most apparent and statistical significance improved considerably (sig.=.01). The most apparent significant difference was between Hispanic males and non- Hispanic males in that 77 percent of Hispanic males were offered literature positions versus 23 percent of their non-Hispanic counterparts. The comparisons for females were significant but less striking, 40 percent of the Hispanic females were offered literature positions versus 26 percent for non-Hispanic females.

As reported in Table 2, there is a weak statistical (sig. <.07), yet substantively important relationship between ethnicity, gender, and whether or not one receives a tenure-track appointment. This weak effect is due primarily to the reduced *n*; there were only 37 respondents in this part of the analysis. While both non-Hispanic males (65 percent) and Hispanic females (65 percent) recorded figures near the overall group average (64 percent), 85 percent of Hispanic males reported tenure-track positions, while only 52 percent of non-Hispanic females received tenure-track appointments. These findings are in line with what Rosenblum and Rosenblum (1990) found to be characteristic of segmented labor markets in academia, namely that those outside the tenure stream or internal labor market are more likely to be women and less likely to be cosmopolitan (p. 158). (Note 3) They also have a greater than 80 percent probability of staying in the non-tenure-track or external academic labor market (Rosenblum and Rosenblum, 1996, p. 441).

Discussion

As noted above, a statistically significant proportion of jobs in the field of literature go to Hispanics, while non- Hispanics are hired primarily to teach language. Said division of labor is particularly observable in the case of males. This pattern becomes meaningful when we consider the relative prestige associated with literature and the concomitant lack of status that comes with an appointment in language. In a recent issue of *Profession*, Russell A. Berman (1997, p. 63) explains:

> While there may be exceptions, language instruction is not as well rewarded with prestige or remuneration as is literary scholarship.... A line of class division, corresponding precisely to the distribution of rewards by the university, runs through all our departments.

Not only are language teaching positions less distinguished, but there are fewer of them, as many of these positions are filled by teaching assistants and part-timers. According to Huber (1996, p. 98), 15 percent of the jobs advertised in the October 1993 "Foreign Language" listings in the Modern Language Association Job List "referred to language teaching expertise only," while "32 percent of the descriptions mentioned expertise in literature only."

One of the rewards of the literature track is the opportunity to teach upper-division literature classes as a graduate student, while simultaneously avoiding the "educational service work" (MLA Final Report 1997) of teaching lower-division introductory language classes. Teaching literature classes, which are generally 3-credit classes, instead of lower-division language classes, which are normally 4-credit classes with high enrollment caps populated with non-majors seeking only to fulfil an exit requirement, translates to a lighter work load, higher pay per hour of instruction, more free time to pursue one's own studies, and a competitive advantage in teaching experience with respect to the job search. (Note 4)

The survey found that non-Hispanics, especially males, were disproportionately relegated to lower-division language classes (mean number of lower division classes taught, Hispanic males vs. non-Hispanic males, was 10.3 and 15.9 respectively; $p<.05$) while serving as teaching assistants. Thus, not surprisingly, the segmentation seen in faculty hiring practices has antecedents in graduate school employment protocols.

Finally, at the faculty level, language teaching positions pay less, a concrete manifestation of the distribution of rewards Berman refers to above. The average entry-level language teaching salary reported was $32,857, while the average entry- level literature position paid $36,607. The average salary, not the entry-level salary, of an assistant professor in a US Department of Foreign Languages is $35,095 (Wright, 1998, p. 115).

As mentioned above, only 6 percent of the survey cohort are domestic-born Hispanics and, if we correct for Puerto Ricans, who are native speakers and non-immigrants, only 2 percent of the cohort are potentially second-generation US Hispanics. As Guadalupe Valdés, a senior member of Stanford's Spanish Department, points out, for reasons of language and social class, "these [2nd generation] speakers of immigrant languages are often considered undesirable in our departments." (1998, p. 154).

This situation is relevant in terms of affirmative action hiring practices, which were designed to remediate prior discrimination. Said remediation does not appear to be present in US Spanish Department hiring practices, and indeed there seems to be a bias against non-native speakers of Spanish, be they of Hispanic origin or not. Hiring practices appear to reflect a *standard language ideology*, which the sociolinguist Rosina Lippi-Green (1997, p. 64) defines as:

> a bias toward an abstracted, idealized, homogenous spoken language which is imposed and maintained by dominant bloc institutions and which names as its model the written language, but which is drawn primarily from the spoken language of the upper middle class.

The "spoken language of the upper middle class" is in this case the spoken Spanish of upper middle class, foreign-born Hispanics. If we assume a correlation between educational levels and social class, then it can be argued that the Hispanic cohort (and the non-Hispanic as well) come from privileged social origins. Forty-six percent of all Hispanics had at least one parent with a 4-year degree or better, and 63 percent of the foreign-born Hispanics arrived in the US with at least a bachelor's degree.

These educational levels become significant when we compare them with the educational levels of the overall US Hispanic population. Only 10 percent of all Hispanics in the US have 4 or more years of college, while the aggregate figure for the US is 27.1 percent (*Digest of Education Statistics 1997*, 1998, p. 17). Moreover, in the US, Hispanics constitute an estimated 10.7 percent of the population but hold only 2.1 percent of Doctor's degrees and 3.1 percent of Master's degrees (*Digest of Education Statistics 1997*, 1998, p. 17).

The elevated educational level, and concomitant social status, of the Hispanic respondents becomes even more evident when we compare these with educational levels in their countries of origin. For example, only 11 percent of the entire population of Spain has a bachelor's degree (Organization for Economic Cooperation 1994), and in Latin America the distinction is even sharper. In Nicaragua, in 1995, "the illiteracy rate represented approximately one-half the population over the age of ten" (Arnove, 1997, p. 92). A mere 8 percent of the entire professoriate in Mexican higher education has a Ph.D. (Altbach, 1996, p. 322). By way of comparison, in the US, 62 percent of all faculty hold a Ph.D. degree (Altbach, 1996, p. 345).

In short, the foreign-born Hispanics represent the educational elite even in the US, but especially in their countries of origin. In marked contrast to the demographics of the US faculty at large, entry-level faculty in US Spanish Departments are predominately Hispanic, upper middle class, and foreign-born. (Note 5)

The reasons for the segmentation observed in Spanish Department hiring practices are doubtless many and complex. Affirmative action policies are clearly a factor, and Literature or Culture is traditionally the domain of the privileged classes. Additionally, as Gerhard Lenski pointed out, once in place, status groups tend "to make in-group membership a resource in the competition for power and privilege" (p. 400), and Bourdieu's (1988) research on class replication in higher education in France supports this view. Similarly, Foucault has pointed to "societies of discourse" and "doctrinal groups" as those responsible for disciplining the order of discourse: "none shall enter the order of discourse if he does not satisfy certain requirements or if he is not, from the outset, qualified to do so" (p. 120). Lang's (1986) language model of discrimination, in consonance with Lippi-Green (1997) and Valdés (1998), argues those who do not acquire privileged language patterns may be crowded into job markets with high densities of similarly marginalized workers. This is a special case of the more general "crowding theory," which argues that marginalized workers are crowded into a limited set of opportunities, thus flooding the labor market in that specific area and so driving down wages for all who compete in that area (Sorensen 1990). This does seem to be consistent with our results, as language jobs pay less across the ethnic and gender continuum, although within the language division Hispanics continue to be paid more on average than non- Hispanics ($34,000 vs $31,000). Furthermore, two factors that the human capital theory of earnings suggest distort the crowding effect--time spent out of the labor force and time taken to train for jobs--are not relevant variables in this survey and, according to Paula England (1982), are fallacious in any event.

Conclusions

Whatever the causes, some of the relations reported in the survey are clearly incongruent with egalitarian principles, especially those which underlie such important programs as affirmative action and the movement toward equal opportunities for women. From a policy standpoint, we would offer the following suggestions as a way to alleviate the reported inequities. First, is the question of salaries. The fact that both Hispanic and non-Hispanic women fare worse than their male counterparts, both in terms of salary and in terms of tenure- track appointments, needs to be rectified.

Some ameliorization of the pay differential between language and literature positions also appears appropriate. Because language teaching serves the vast majority of students, and not incidentally generates the bulk of tuition dollars, it is in the long-term best interest of those in literature to avoid the inequities which are currently splitting Spanish Departments. As Dorothy James (1997) points out,

> In largely confining language teaching to the lowest levels of the curriculum and in assigning it to a category of teachers different from those who teach literature or other content, we have undermined the rationale for having a paid, full-time professoriat in foreign languages....Can we really be surprised if it dawns on administrators under severe budgetary constraints that they can save a lot of money by doing away with the small upper-level foreign literature programs altogether and sustain the lower-level language courses without the benefit (and expense) of properly paid senior faculty members? (p. 49)

Likewise, literature positions should be made available to more non-Hispanics, given that the "crowding" or competition between non-Hispanics for the limited number of language positions drives down pay scale, and here we can point to the use of graduate students and part-timers as related aggravating factors. This would also add diversity of ideas and perspectives to the discipline.

Another issue is the co-mingling of departmental diversity data within colleges. Under-representation would be better determined on a departmental basis. Allowing administrators to report only aggregate college figures may well bias Spanish Department hiring patterns, and so produce the oft-cited "small worlds" or "barrioization" of and in Spanish Departments (Clark, 1987; Garza, 1988). Said reporting may also influence student recruitment and retention, and later give rise to pipeline deficiencies (Bernhardt, 1995; James, 1989). Appropriate steps need be taken to train, recruit, and retain both Hispanic and non-Hispanic candidates from under-represented ethclasses. (Note 6)

Finally, exchange programs, at both the student and faculty level, should be expanded. Present policy encourages "brain drain" from developing nations at a time when they most need their best and brightest (Altbach, 1996, p. 302), and simultaneously disregards the crisis in the US job market (Curren, 1994; Modern Language Association Committee on Professional Employment, 1997).

We conclude by observing that the assumption that equates Hispanic hires with increased social diversity and mobility can be illusory in some situations. It may be advisable to reassess the ultimate fairness of hiring policies that, while empowering under-represented ethnic groups, do so within a context of socio- economic or class bias. Therefore, in order to fully appraise academic employment practices in the US, we feel it is necessary to consider not only ethnicity and gender, but also national origin and, importantly, how all these intersect with social class.

Notes

1. The questionnaire is based upon the survey instrument used by the National Research Council in assessing earned doctorates in the US in 1995. It was modified to ascertain Spanish Department-specific information.

2. Only two Hispanics indicated they were non-native speakers. All other Hispanics declared themselves native speakers.
3. In Gouldner's (1957) sense of the term, a cosmopolitan is one with high professional skills (native speakership), low loyalty to his or her employing institution, and high allegiance to a reference group of similar (socio-ethnically as well as intellectually) academics located outside the individual's institution.
4. The teaching-load disparity persists into the arena of entry level jobs. 54% of Hispanic males and 45% of Hispanic females report teaching-loads of 5 classes per year or less, while only 16% of non-Hispanic males and 26% of non-Hispanic females teach 5 or fewer classes per year (sig. <.07). Overall, 31% of the survey cohort teach 5 or fewer classes per year.
5. Overall, the US faculty in higher education is 2.5% Hispanic, 87% white not- Hispanic, 13% foreign-born, and 68% male (Finkelstein, Seal, and Schuster, 1998: 28).
6. Milton M. Gordon defines "ethclass" as the "social space created by the intersection of the ethnic group with the social class" (51).
7. All data regarding the 1996 population is from the National Opinion Research Center (1997).
8. Job placement figures are taken from the MLA's survey of PhD placement (Huber 1996), the most recent year for which figures are available. Data are only collected every two years, and the 1996 data are no yet available. The survey does not break out tenure by ethnicity, nor does it address pay or the language/literature distinction.

References

Altbach, Phillip G. (Ed.) (1996). *The International Academic Profession: Portraits of Fourteen Countries*. Princeton: The Carnegie Foundation for the Advancement of Teaching.

Arnove, Robert E. (1997). Neoliberal Education Policies in Latin America. In Carlos Alberto Torres and Adriana Puiggrós (Eds.), *Latin American Education: Comparative Perspectives* (79-100). Boulder, CO: Westview Press.

Berman, Russell A. (1997). Reform and Continuity: Graduate Education Toward a Foreign Cultural Literacy. *Profession 1997*, 61-74.

Bernhardt, Elizabeth B. (1995). Teaching Literature or Teaching Students? *ADFL Bulletin*, 26.2, 5-6.

Bourdieu, Pierre (1988). *Homo Academicus*. Trans. Peter Collier. Stanford: Stanford University Press.

Clark, Burton R. (1987). *The Academic Life: Small Worlds, Different Worlds*. Princeton: Princeton University Press.

Curren, Eric D. (1994). No Openings at This Time: Job Market Collapse and Graduate Education. *Profession 94*, 57-61.

England, Paula (1982). The Failure of Human Capital Theory to Explain Occupational Sex Segregation. *The Journal of Human Resources*, 17.3, 359-70.

Eutis, Christopher (1997). Dissertations 1996. *Hispania*, 80.2, 388-403.

Finkelstein, Martin J., Robert K. Seal, and Jack H. Schuster (1998). *The Academic Generation: A Profession in Transformation*. Baltimore, London: Johns Hopkins.

Foucault, Michel (1984). The Order of Discourse. In M. Shapiro (Ed.), *Language and Politics* (pp. 108-138). New York: NYUP.

Fuss, Diana (1989). *Essentially Speaking*. New York, London: Routledge.

Garza, Hisauro (1988). The "Barrioization" of Hispanic Faculty. *Educational Record*, 68.4, 122-124.

Gordon, Milton M. (1964). *Assimilation in American Life*. New York: Oxford UP.

Gouldner, Alvin W. (1957). Cosmopolitans and Locals: Towards an Analysis of Latent Social Roles-I. *Administrative Science Quarterly*, 2, 281-306.

Huber, Bettina J. (1994). Recent Trends in the Modern Language Job Market. *Profession 94*, 87-105.

Huber, Bettina J. (1996). The MLA's 1993-94 Survey of PhD Placement: The Latest Foreign Language Findings and Trends Through Time. *ADFL Bulletin*, 27.3, 58-77.

James, Dorothy (1989). Reshaping the "College-Level" Curriculum: Problems and Possibilities. In Helen S. Lepke (Ed.), *Shaping the Future: Challenges and Opportunities* (pp. 79-100). Lincolnwood, Illinois: Natl. Textbook.

James, Dorothy (1997). Bypassing the Traditional Leadership: Who's Minding the Store? *Profession 1997*, 41-53.

Anon. (1998). Job Tracks: Who Got Hired Where, 1997-1998. *Lingua Franca*, 6.1, 57-86.

Kramsch, Claire (1997). The Privilege of the Nonnative Speaker. *PMLA*, 112.3, 359-369.

Lang, Kevin (1986). A Language Theory of Discrimination. *Quarterly Journal of Economics*, 101.2, 363-382.

Lederer, Herbert (1980). The "Native Speaker" Issue: Problem or Pretext? ERIC database, no. EJ253826.

Lenski, Gerhard E. (1966). *Power and Privilege: A Theory of Social Stratification*. New York: McGraw-Hill.

Lippi-Green, Rosina (1997). *English With an Accent: Language, Ideology, and Discrimination in the United States*. London and New York: Routledge.

Modern Language Association Committee on Professional Employment (1997). *A Guide: Evaluating the Mission, Size, and Composition of Your Doctoral Programs*. New York: Modern Language Association.

Modern Language Association Committee on Professional Employment (1997). *Final Report*. New York: Modern Language Association.

National Center for Education Statistics (1998). *Digest of Education Statistics 1997*. Washington, DC: United States Department of Education.

National Opinion Research Center (1997). Affirmative Action Table #1: Preliminary Number of PH.D.S [sic], By Fine Field of Doctorate, Race/Ethnicity, Citizenship, and Gender, 1996. Chicago: NORC.

Organization for Economic Cooperation and Development (1994). *Education at a Glance*. Paris: Organization for Economic Cooperation and Development.

Paikeday, Thomas M. (1985). *The Native Speaker is Dead!*. New York: PPI.

Rosenblum, Gerald, and Barbara Ruben Rosenblum (1990). Segmented Labor Markets in Institutions of Higher Learning. *Sociology of Education*, 63, 151-164.

Rosenblum, Gerald, and Barbara Ruben Rosenblum (1996). The Flow of Instructors Through the Segmented Labor Markets of Academe. *Higher Education*, 31, 429-445.

Sorensen, Elaine (1990). The Crowding Hypothesis and Comparable Worth. *The Journal of Human Resources*, 25.1, 55-89.

Valdés, Guadalupe (1998). The Construct of the Near-Native Speaker in the Foreign Language Profession: Perspectives on Ideologies about

Language. *Profession 1998*, 151-160.

Wright, John W. (Ed.) (1998). *The American Almanac of Jobs and Salaries: 1997-1998 Edition*. New York: Avon.

About the Authors

Jerry Hoeg

Jerry Hoeg is Associate Professor of Spanish at Pennsylvania State University. His email address is jvh1@psu.edu

Eric Cohen

Eric Cohen is an Assistant Professor of Sociology at Penn State.

Christine Fullen

Christine Fullen was formerly in the Department of Health and Human Services at Penn State. She now is employed by Fayette County in the Commonwealth of Pennsylvania.

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs- Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
Arizona State University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa- DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

```
PCL XL error

        Subsystem:  KERNEL

        Error:      IllegalOperatorSequence

        Operator:   SetClipRectangle

        Position:   13
```

BEST COPY AVAILABLE

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 787 times since April 11, 1999

Education Policy Analysis Archives

Volume 7 Number 13 April 11, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

BEST COPY AVAILABLE

Contextualizing Homeschooling Data: A Response to Rudner

Kariane Mari Welner
University of California, Los Angeles

Kevin G. Welner
University of Pennsylvania

Abstract

Rudner (1999) presents the results of a survey and testing program, administered by Bob Jones University (BJU), for homeschooling students. In this response, we applaud Rudner's contribution to building a greater understanding of the homeschooling movement. However, we also voice a strong concern that what Rudner contributed with one hand, he took back with the other. We contend that Rudner's analysis of the BJU data fails to offer a straightforward explanation of important and striking limitations. The unfortunate result is an inaccurate portrayal of homeschoolers as a white, Christian, monolithic population. Although the results of Rudner's analyses are likely valid for the particular population he studied, his insufficient attention to the data's bias has led to an erroneous picture of homeschooling.

At a time when most educational researchers pay little or no attention to the homeschooling movement, we appreciate Dr. Rudner's (1999) work as well as the decision of *EPAA* to give that work a platform. Although Rudner's article does not vary far (either in methodology or in findings) from much of the body of the homeschooling research already in print (e.g., Gustavsen, 1981; Mayberry, 1987; Ray, 1990, 1997; Wartes, 1988), keeping this research current and expanding the sample size contributes greatly to the limited assortment of currently-published studies. We believe that the questions asked of the data in this study are important; these issues need to be further explored. However, the data that Rudner analyzed are derived from only one section of the homeschooling population. And here lies the article's weakness: it fails to explain this limitation in a way that adequately alerts readers.

While we do not disagree with Rudner's tentative conclusions concerning homeschoolers' performance on standardized tests, we do think there is a need to offer several cautions to the readers of this study. The data employed in this study were taken from parents who used the Bob Jones University (BJU) standardized testing program. Rudner's article only briefly, and inadequately, addresses the fact that this may not be a cross-section of the homeschooling population: "...it should be noted that it was not possible within the parameters of this study to evaluate whether this sample is truly representative of the entire population of home school students" (this quotation is from the article's "Discussion" section). Rudner does not explain the relevance of this potential limitation as regards the demographic and achievement information that constitute the heart of these analyses. Nor does he offer the obvious reason why the BJU data may not be representative of the larger population: The University's image, at least partially deserved, is of racial intolerance and religious orthodoxy. Accordingly, some of Rudner's conclusions (e.g., that homeschoolers are overwhelmingly white and Christian) should instead be read as limitations on some of his other conclusions (concerning, e.g., median income, marital status, and achievement levels on standardized tests).

A related caution, which we would have liked Dr. Rudner to have offered, is that the data base drew a non-random, two- percent sample (even by the most conservative estimates) of the homeschooling population. (Note 1) Given that Rudner's sample involves 20,760 students, this sample then constitutes anywhere from 1.28% to 2.08% of the homeschooling population. Because the sample was biased in favor of a population associated with BJU, extrapolations from that data are very unreliable.

Yet Dr. Rudner, in his abstract, states that his article seeks to answer the following questions:

> Does home schooling tend to work for those who chose to make such a commitment? That is, are the achievement levels of home school students comparable to those of public school students? Who is engaged in home schooling? That is, how does the home school population differ from the general United States population?

Notwithstanding these broadly worded questions, we note that Rudner acknowledges the fact that his data are not derived from "a controlled experiment" and must be understood within that context. Yet his acknowledgment fails to detail the broader context—namely those issues associated with BJU. He also fails, when setting forth these questions that he seeks to answer with this data, to heed his own warning about its limitations. Contrary to his stated aim, this data simply cannot be used to reliably compare homeschoolers' achievement levels with those of the general population or to describe the demographics of homeschoolers.

In addition to our concerns about the generalizability of Rudner's conclusions, we are concerned that Rudner's relative neglect of issues surrounding the selected sample will serve to perpetuate the common view of homeschoolers as a narrow and easily-defined section of the American population. Because of his omission, only those readers who already know about the full scope of homeschoolers or about BJU's image could raise this red flag. Importantly, the media's coverage of the release of Rudner's study portrayed his conclusions as descriptive of the broader homeschooling population (see Archer, 1999; Cook, 1999; Mathews, 1999; Schnaiberg, 1999; Toomer-Cook, 1999). We found only one article that even mentioned, albeit briefly, the ill fit between the homeschooling population described by Rudner and the homeschooling population in the area (south Florida) served by the newpaper (Nazareno, 1999).

Homeschoolers' Diversity

Notwithstanding the picture painted by the data presented in Rudner's article, today's homeschooling families represent a diverse sampling of the American population. Although many homeschoolers remain white and middle-class, the recent upsurge in homeschooling has drawn people from all ethnic and class groups (Knowles, 1988; Nazareno, 1999; Wahisi, 1995). Ideologically, parents who homeschool represent a similarly broad cross-section of American society (Knowles, 1988). While in its recent resurgence homeschooling began as a trend among fundamentalist Christians with primarily religious motivations, homeschoolers now represent a wide array of values and political mores (Bolick, 1987; Mayberry, 1987; Van Galen, 1988). As such, homeschoolers are no longer an easily defined segment of the population. (Note 2) Rudner's article, therefore, would have been more comprehensive and accurate had he acknowledged existing research demonstrating that his sample was not representative of the broader population.

Not All Homeschoolers Give Tests

In contrast to what is presented in Rudner's data, there exists a large and growing portion of the homeschooling population that does not administer standardized tests to its children. (Note 3) While some homeschoolers employ the "school at home" methodology that Rudner's questions allude to, replete with curricula and testing, other homeschoolers avoid these practices. These parents, often referring to themselves as "unschoolers," follow the philosophy of the late John Holt (see Holt 1981; 1983; 1989). They often choose to homeschool in order to avoid what they view as the restrictiveness of set curricula and testing (Franzosa, 1991). They believe in allowing a child's natural curiosities to set the scope and pace of education, even if it means waiting a long time before the child expresses interest in a particular topic (Wartes, 1988a). Feeling that the manner in which schools teach is not the way that children learn, they often view standardized testing as a part of the misguided system that they have left behind, and they put great effort into avoiding such testing in their children's education (Common & MacMullen, 1986; Gibbs, 1994). Notwithstanding the important role that this segment of the homeschooling population plays within the larger movement, its existence is not noted in Rudner's article.

Bob Jones University

The data Rudner analyzed was derived exclusively from parents who used the testing services of Bob Jones University, a fundamentalist Christian institution located in Greenville, South

Carolina. The university prohibits interracial dating and marriage between its African-American and white students. Prior to 1971, African Americans were banned outright from attending the university (White, 1982). These racial policies were the subject of highly-publicized litigation before the United States Supreme Court in 1983, concerning the question of whether the University, given its explicit racially discriminatory policies, could maintain its "501(c)(3)" tax-exempt status (Bob Jones University v. United States, 1983; White, 1983).

Importantly, the University's racial views are anathema to many Americans, whether they be Christian or non-Christian, fundamentalist or non-fundamentalist. While many parents using the BJU testing service may not share the University's convictions, many other families have no doubt chosen not to employ the services of BJU precisely because of BJU's racial stance. The racial distribution of Rudner's homeschooled students showed 0.8% African American and 0.2% Hispanic. These statistics become much more meaningful when grounded in an understanding of the data source.

BJU's religious orientation may also have prompted many homeschoolers to shun a relationship with the University's testing service. The BJU web pages trumpet its Biblical grounding, noting that the University is "both orthodox and fervent in its evangelistic spirit" (see the Bob Jones University Website, http://www.bju.edu/aboutbju/history/). The religious distribution of Rudner's homeschooled students showed almost 58% Independent Fundamentalist, Baptist, or Independent Charismatic and only 6% non-Christian. These statistics, too, become much more meaningful when grounded in an understanding of the data source.

The Article's Perspective

Dr. Rudner is an accomplished scholar in the field of assessment. Using this expertise, he has testified on behalf of homeschoolers represented by the funder of this study, the Home School Legal Defense Association (HSLDA). We wish to acknowledge the contributions of the HSLDA in blazing a trail for the legal rights of homeschoolers. However, it is also important to note that the HSLDA is an advocacy organization. Moreover, while the HSLDA does not exclude from membership those who hold non-Christian beliefs, it is an overtly Christian establishment with an agenda dedicated to supporting the rights and duties of families as commanded by Biblical mandate.

We recognize that Dr. Rudner may have little or no experience with the homeschooling population outside of this Christian context and therefore might not be aware that many homeschoolers do not fit within the segment of the population who would consider employing the services of BJU. Like many Americans, Dr. Rudner may have simply taken for granted that the homeschooling population remains very narrow. When an analysis is inaccurately premised on the assumption that homeschooling is a phenomenon that is almost exclusively limited to conservative Christian parents, there is less reason to question the representativeness of a sample drawn from BJU.

In understanding this article's perspective, we also note that the HSLDA funded this study, at least in part, as a vehicle for gaining political support on behalf of homeschoolers. Toward this end, the organization distributed copies of the study to members of Congress. Michael Farris, the president of HSLDA, explained that he "hope[s] that [the study] will help judges and public policy makers make better decisions about our freedom" (Billups, 1999). We share this goal as well the belief that legislation concerning homeschooling is best based on a complete understanding of the homeschooling population. Accordingly, the following section presents a reconsideration of some of Dr. Rudner's findings in light of the source of his database.

The Article's Conclusions

As a general matter, we suggest that the data from Bob Jones University add to our understanding of the particular homeschooling population served by BJU and raise interesting questions about the broader homeschooling population. But we would stop far short of drawing the more universal conclusions trumpeted by the HSLDA.

Consider, for example, Dr. Rudner's analysis of demographic information, concerning the high levels of formal education obtained by homeschooling parents and the high median income of homeschooling families. We caution against drawing firm conclusions from this data, particularly the analysis concerning median family income. Not all states require children to take standardized tests (and those that do generally provide a way for parents to take the tests without cost to the family). Consequently, those families who elect to pay a testing service may be in a higher income bracket than those who do not. Moreover, a 1990 survey of Maine homeschoolers revealed that 70% of respondents had an annual pretax income of less than $35,000 (Lyman, 1993). While this Maine study also had many limitations, it nonetheless raises the question of the generalizability of Rudner's findings. The analyses concerning homeschooling parents' level of formal education, computer use by students, and the amount homeschooling parents spend on school supplies could be tied to this income data and influenced by these same factors. Also, with regard to expenditure on school supplies, additional issues are raised (e.g., whether the parents have supplies from an earlier child and whether they can borrow supplies from a friend or family member).

Rudner also found that 98% of homeschooling students in the BJU data base live within married couple families. This statistic should, we contend, have been presented within a context explaining the conservative nature of BJU and its view of divorce as unbiblical. Further, given the view held by most conservative Christians that a woman's primary commitment is to her husband and her children, Rudner conclusion that 77% of homeschool mothers in this data base do not participate in the labor force is, we believe, also better understood within the BJU context. (Note 4)

Rudner notes that homeschooled students watch much less television than do most students nationwide (with 65% of homeschooled students in the BJU data base watching one hour or less per day compared to 25% nationally). But again, conservative Christians tend to have strong moral objections to the quality of television programming—much more so than the general population. Accordingly, it may be a certain set of moral standards rather than homeschooling that drives this result.

The article further states that the "primary focus of many home schools is on religious and moral values." But, while many homeschoolers do see this as their primary focus and purpose, many do not—recall the "unschoolers" described above. Van Galen (1988) describes a group of parents, whom she labels as "Pedagogues," whose motivation is decidedly secular. These parents:

> ... teach their children at home primarily for pedagogical reasons. Their criticisms of the schools are not so much that the schools teach heresy, but that the schools teach whatever they teach ineptly.... While diverse in other aspects of their lives, they share a respect for their children's intellect and creativity and a belief that children learn best when pedagogy taps into the child's innate desire to learn (p. 55).

While these parents may have religious beliefs, the reason they chose to homeschool was not religiously motivated, and the focus of these homeschools is not the teaching of "religious and moral values."

Other demographic characteristics that Rudner ascribes to homeschoolers could also be a result of the population sampled, a possible conflation of almost his exclusively Christian population with the trait of homeschooling. For example, it may be simply that Christian families are larger than those in the general population, not necessarily that most homeschooling parents have more children. Likewise, the fact that almost one in four homeschooled students in this study have a certified teacher as a parent may also be tied to the overwhelmingly Christian population, since many young Christian women see teaching as one of the few appropriate areas of employment and as a field where they can develop skills useful in both family and church life. Finally, as mentioned previously, the fact that the demographics reported by Rudner showed an almost exclusively white, Christian population could also be an artifact of the data source.

As Dr. Rudner noted, family income is strongly correlated with children's test scores. However, we do not know whether the general population of homeschoolers has the same high level of income as the families in the BJU data base. Further research is needed to demonstrate whether or not this difference in test scores would hold up if a lower income sample of homeschoolers were tested. That said, we do believe that homeschooled students can attain a significant benefit from the one-on-one learning experience, and this could be a powerful factor in driving higher test scores. (Note 5)

Rudner concluded by stating that "these comparisons between home school students and students nationwide must be interpreted with a great deal of caution," and that "the reported achievement differences between groups do not control for background differences in the home school and general United States population and, more importantly, cannot be attributed to the type of school a child attends." Some researchers, in fact, would say that the test scores have nothing to do with how the children were schooled and simply show the results expected for children that come from this demographic group—households that are overwhelmingly white, well educated, two-parent, and middle class (see Coleman et al., 1966; Ogbu, 1987). This is not to say that these parents did not do a good job teaching their children, it is only to say that a comparable sample within the public or private schools may have scored just as well.

Our Conclusions

The actual analyses conducted by Dr. Rudner are important. Our critique is offered as a cautionary supplement, rather than as an objection, to his contribution. We feel that a more thorough explanation of the data's source and context helps us to build a better understanding of America's homeschooling population.

Notes

1. Patricia Lines conservatively estimates the number of homeschooled children at approximately 1 million (Lines, 1998). Less conservative appraisals among homeschooling associations and researchers place the number of homeschooled students at more than 1.2 million students (Hawkins, 1996; Kennedy, 1995; Ray, 1997). Newsweek recently estimated that number at 1.5 million (Kantrowitz & Wingert, 1998), a figure that the Home School Legal Defense Association —the sponsor of Rudner's study—also circulates in its literature. Others estimate that the number is as high as 1.6 million (see Yarnall, 1998).
2. There are also many resources directed at homeschoolers that are not characterized by the demographics provided by Rudner (see the following web pages: "Bnos Henya Project: Jewish Orthodox Homeschooling"; "Al-Madrasah Al-Ula: The Magazine for Muslim Home Schoolers"; "Native American Homeschool Association Web Site"; and "Pagan Homeschool Page").
3. Of course, if they are required to do so by law, then they comply with the state's requirements (these requirements for homeschoolers vary from state to state). However, these parent engage in the testing merely to satisfy their legal obligation, not because they believe testing to be an educationally worthwhile practice.
4. It is true that generally, in order to homeschool, one of the parents must possess the ability to remain at home throughout the day, thus allowing that parent to teach and supervise the children. However, homeschooling parents may have jobs permitting them to also supervise their children —either through a flexible schedule, a home-based business, or a job allowing for on-site

supervision of their children—with the result that both parents become part of the labor force.

5. Further, many important forms of knowledge, which homeschooling parents may emphasize in their children's education, may not be assessable by standardized tests.

References

Hyperlinks to some of these documents

Al-Madrasah Al-Ula: The Magazine for Muslim Home Schoolers. [online].

Apple, M. (1982). *Education and power*. London: Routledge & Kegan Paul.

Archer, S. (1999, March 24). New evidence supports home schooling: Students perform better than those in classrooms. WorldNetDaily. [online]

Billups, A. (1999, March 24). Students in home schools perform better on tests. *The Washington Times*, p. A8. Bnos Henya Project: Jewish Orthodox Homeschooling. [online]

Bob Jones University v. United States. (1983). 461 U.S. 574. [online]

Bob Jones University Website. [online].

Bolick, C. (1987). The home-schooling movement. *The Freeman: Ideas on Liberty*, 37(3), 84-89.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., MacPartland, J., Mood, A. M., Winfeld, F., and York, R. (1966) *Equality and educational opportunity*. Washington, DC, Office of Education, US Government Printing Office.

Common, R. W. & MacMullen, M. (1986). Home schooling...a growing movement. *Education Canada*, 26(2), 4-7.

Cook, S. (1999, March 25). "Report card on home schooling in US: Study finds children taught by parents perform above national average." *The Christian Science Monitor*. [online]

Franzosa, S. (1991). The best and wisest parent: A critique of John Holt's philosophy of education. In J. Van Galen's (Ed.) *Home Schooling; Political, Historical, and Pedagogical Perspectives* (pp.121-135). Norwood, NJ: Ablex Publishing.

Gibbs, N. (1994). Home sweet school: Seeking excellence, isolation, or just extra "family time," more and more parents are doing the teaching themselves. *Time*, 144(18), 62-63.

Gray, S. (1992). *Why some parents choose to home school.* Dissertation. Los Angeles, CA: UCLA.

Greene, S. (1984). Home study in Alaska: A profile of K-12 students in the Alaska centralized correspondence study program. *ERIC Document Reproduction Service*, 255(494).

Gustavsen, G. (1981). *Selected characteristics of home schools and parents who operate them*. Doctoral Dissertation, Andrews University. University Microfilms International No. 8205794.

Hawkins, D. (1996). Homeschool battles: Clashes grow as some in the movement seek access to public schools. *U.S. News & World Report*, 120(6), 57

Holt, J. 1981. *Teach Your Own: A Hopeful Path for Education*. New York, NY: Delta/Seymour Lawrence.

Holt, J. 1983. *How Children Learn*(revised ed.) New York, NY: Delta/Seymour Lawrence.

Holt, J. 1989. *Learning All the Time*. Reading, MA: Addison-Wesley Publishing.

Kantrowitz, B. & Wingert, P. (1998, October 5). Learning at home: Does it pass the test? *Newsweek*, n. 40, p. 64.

Kennedy, J. G. (1995). Home schooling grows up: Teaching at home moves to the cutting edge of education. *Christianity Today*, 39(8), 50.

Knowles, J. G. (1988). Introduction: The context of home- schooling in the United States. *Education and Urban Society*, 21(1), 5-15.

Lines, P. (1987). An overview of home instruction. *Phi Delta Kappan*, 68(7), 510-517.

Lines, P. (1998, Spring). *Homeschoolers: estimating numbers and growth*. National Institute on Student Achievement, Curriculum, and Assessment, Office of Education Research and Improvement: U.S. Department of Education.

Lyman, I. (1993). Better off at home? *National Review*, 45(18), 62-63.

Mahan, B. M. & Ware, B. J. (1987). *Home schooling: Reasons parents choose this alternative form of education and a study of attitudes of home-schooling parents and public school superintendents toward the benefits of home- schooling*. Master's Project, University of Dayton. ERIC Document No. ED286624.

Mathews, J. (1999, March 24). A home run for home schooling:

BEST COPY AVAILABLE

Movement can point to high test scores in national study. *The Washington Post*, p. A11. [online].

Mayberry, M. (1987). The 1987-1988 Oregon home school survey: And overview of findings. *Home School Researcher*, 4(1), 1-9.

Native American Homeschool Association Web Site: Saving Our Culture For Our Children Through Our Children. [online].

Nazareno, A. (1999, March 24). Home schools effective, group says after study. *The Miami Herald*.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven: Yale University Press.

Ogbu, J. U. (1987). Variability in minority school performance: A problem in search of an explanation. *Anthropology and Education Quarterly*, 18(4), 312-334.

Pagan Homeschool Page. [online]. Available at: http://members.aol.com/Barbooch/index.html.

Ray, B. (1990). Home schools: A synthesis of research on characteristics and learner outcomes. *Education and Urban Society*, 21(1), 16-31.

Ray, B. (1997). *Strengths of their own: Academic achievement, family characteristics, and longitudinal traits*. Salem, OR: National Home Education Research Institute Publications.

Rudner, L. M. (1999). Scholastic achievement and demographic characteristics of home school students in 1998. *Education Policy Analysis Archives*, 7(8). [online]. Available at http://epaa.asu.edu/epaa/v7n8/.

Schnaiberg, L. (1999, March 31). Study finds home schoolers are top achievers on tests. *Education Week*, 18(29), 5. [online] .

Shepherd, M. S. (1986). The home schooling movement: An emerging conflict in American education [abstract]. *Home School Researcher*, 2(1), 1.

Toomer-Cook, J. (1999, March 24). Home schoolers are making the grade, national study says. *Deseretnews.com*. [online].

Van Galen, J. (1987). Explaining home education: Parents accounts of their decisions to teach their own children. *The Urban Review*, 19(3), 161-177.

Van Galen, J. (1988). Ideology, curriculum, pedagogy in home education. *Education and Urban Society*, 21(1), 52- 68.

Wahisi, T. T. (1995, October). Making the grade: Black families see the benefits in home-schooling. *Crisis*, 102(7), 14-15.

Wartes, J. (1988). The Washington home-school project: Quantitative measures for informing policy decisions. *Education and Urban Society*, 21(1), 42-51.

White, E. (1982, October 20). Racial policies, religious rights square off as high court hears tax-exemption case. *Education Week*. [online].

White, E. (1983, June 1). Court bars tax breaks for discriminatory schools. *Education Week*. [online].

Williams, D. D., Arnoldson, L. M., & Reynolds, P. (1984). Understanding home education: Case studies of home schools. Paper presented at the annual meeting of the American Education Research Association, New Orleans.

Yarnall, L. (1998, October 29). Online businesses tap home-schooling market. *New York Times*, v. 148, p. D1.

About the Authors

Kariane Mari Welner

University of California, Los Angeles
Graduate School of Education & Information Studies
405 Hilgard Avenue
Los Angeles, CA 90095-1521

Email:knemer@ucla.edu

Ph.D. UCLA, in progress
M.A. UCLA, 1996
B.A. UCLA, 1995

Kariane Mari Welner is a doctoral candidate at UCLA. Her specializations include homeschooling and issues of sociology and democracy in education.

Kevin G. Welner

University of Pennsylvania
Graduate School of Education
3700 Walnut Street
Philadelphia, PA 19104-6216
(215) 898-5355

kevinw@gse.upenn.edu

Ph.D. UCLA, 1997
J.D. UCLA, 1988
B.A. UC Santa Barbara, 1985

Kevin G. Welner is a visiting researcher and lecturer at the University of Pennsylvania. His specializations include program evaluation and educational policy issues, particularly those concerning organizational change and school law. His personal homepage is at http://www.gse.upenn.edu/welner/ .

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa- DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA
• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 561 times since April 13, 1999

Education Policy Analysis Archives

Volume 7 Number 14 April 13, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Supporting Community-Oriented Educational Change: Case and Analysis

Linda Mabry
Indiana University

Laura Ettinger
Indiana University

Abstract

A study of a federally funded program to develop and implement community-oriented social studies curricula and curriculum-based assessments grounds cautions for educational change initiatives. In this case, despite the project director's stated intent to support teachers' desire for instruction regarding local culture and history, top-down support for classroom-level change evidenced insensitivity. Production and implementation of the planned curricula and assessments was obstructed by teacher's lack of cultural identification with the targeted community groups, workload, competing instructional priorities, inadequate communication, and organizational politics. Professional development was sometimes beneficial but more often ineffective—either perfunctory, unnecessary, or disregarded. The findings offer insight unnecessary, or disregarded. The findings offer insight regarding educational change and a systemic analysis.

An evaluation study of a federally funded program offered an appealing opportunity to study an instance of community-oriented educational change in a project to develop and implement social studies curricula related to the local history and cultural traditions of school children in three regions of the country (Note 1). Anticipating that the study would provide empirical support for conceptualizations of successful school reform as emanating from local decision-makers and implementers, we were surprised to learn instead some sobering lessons about limitations inherent in policies and practices intended to support school-based education efforts emphasizing communities.

Top-down reform has been criticized as insensitive and unrealistic, but how should policy-makers and funders encourage and support communities and their educators? After a description of the project and the evaluation, findings will be organized according to emergent issues and then by Bronfenbrenner's (1979) levels of ecological analysis. It is not the notions of community-oriented education or teacher-designed reform which cause concern but a complex of implementation motives and issues. This case study is offered to deepen through vicarious experience the understanding of educators and others interested in sophisticating efforts to improve education.

The Heritage Project

The Heritage Project was a three-year federally sponsored program to develop and implement social studies curricula based on local history and cultural traditions in rural public schools in three culturally diverse regions of the country. The stated intention of the project director, a university professor, was to support practitioners in developing and implementing community- oriented social studies curricula. The idea resonated with Fullan's (1991) and Sarason's (1990) contention that fundamental educational change requires the involvement of practitioners and with the groundswell for teacher empowerment (see Astuto, Clark, Read, McGree, & deKoven Pelton Fernandez, 1994).

At each school, Heritage Project teachers were to create social studies curricula focused on local communities in year 1, implement curricula in year 2, and develop and implement curriculum-based assessments in year 3. The project director and a central office coordinator were to support these efforts by arranging professional development, facilitating networking among sites, and distributing funds for materials and other needs. A local site coordinator in each state was to ensure smooth interface, focus on program objectives, and provide assistance to teachers. At historically and culturally distinct sites in the Southeast, Midwest, and Southwest, fourteen teachers and approximately 200 students participated. Each site was unique and complex. Contextual issues at each site played a major role in the life and success of the project.

Table 1
The Heritage Project: Sites and Personnel

Heritage Project Central Office
Project director—university professor
Central office coordinator—graduate student

Midwest site	**Southeast site**	**Southwest site**
Site coordinator:	*Site coordinator*:	*Site coordinator*:
yr. 1: graduate student yr. 2-3: central office coordinator	college professor	university professor
School 1	**School 1**	**School 1**
Caucasian	African American	Hispanic
(3 teachers*) (43 students)	(2 teachers) (29 students)	(1 teacher*) (36 students)
School 2	**School 2**	**School 2**
Caucasian	African American	Native American
(4 teachers*) (10 students)	(1 teacher) (30 students)	(2 teachers) (25 students)
	School 3	
	mixed ethnicity	
	(1 teacher) (30 students)	
Schools were in 2 school districts.	All schools were in 1 school district.	Both schools were in 1 school district.

* Number of teacher-participants fluctuated from year to year. The number listed indicates those who were participating at the end of the project. Numbers of student-participants fluctuated from year to year at all schools.

Midwest site

In the Midwest, where the project was implemented in two elementary schools in non-adjacent school districts, significant issues regarding teacher ideology and autonomy clouded progress. The lead teachers at each school, former students of the project director, were confident of their training and experience. Their self-directedness in the face of historically high principal turn-over in rural schools helped assure program continuation in an environment of fluctuating attention and support. But because of their confidence and autonomy, they felt free to follow their own ideas, including those which conflicted with program goals. Although they had expressed eagerness regarding community-oriented learning, they chose to thwart the program by merely extending their existing social studies curricula rather than developing new curricula emphasizing local history and culture. Their resistance to community-oriented curricula manifested conflict between their beliefs about their communities as cultural microcosms appropriate instructional goals and project philosophy. Although they preferred rural life, these teachers viewed their all-white communities as threads in the national fabric, not culturally or historically distinct. They believed it more important to introduce their students to the wider world than to the local history and culture of their remote communities. Teacher acceptance of the foundational principles of the program did not occur until, at the close of year 2, they bent to pressure from the program central office.

It was not quite a game of musical chairs, but there were a number of personnel changes at the Midwest site in the three-year life of the project. Some were initiated by the project director in response to teacher unwillingness to work toward project goals. Others were initiated locally as teachers objected to unexpected project workloads. The site coordinator left the program at the end of year 1, and the unpopularity of the insistent central office coordinator who doubled as her replacement added irritation to discontinuity. At the conclusion of the project, only two teachers remained of the original personnel. Although the project director eventually touted the development of working relationships among these teachers as an important project achievement, in fact personnel changes diverted attention from project goals and undercut the development of working relationships, as new people constantly needed orientation.

Southeast site

Three schools in one district were involved in the Southeast. Formidable difficulties arose because the Heritage Project teachers were not self-selected but reluctant participants in the program, identified by a school district coordinator who intended to use the program as a means to upgrade their skills. Luckily, two teachers were members of the local African American community whose culture and history were the focus of the program at that site. But a third teacher resisted inclusion of African American culture in the curriculum, and a fourth resisted the considerable extra effort required in curriculum development. Professional development and the urging of the site coordinator eventually led to a shift in favor of teaching local history, and some teachers eventually described their participation in the program as transformative. But during much of the three years of the Heritage Project, half of the teachers exhibited strong passive resistance.

The site coordinator complained that project expectations were out of line with actual possibilities and cited the director's lack of appreciation for local culture, expectations, relationships, and working styles as the heart of the problem. Urging the teachers toward project goals, she exercised a strong management style ultimately pivotal in producing what the project director called the strongest curriculum in the three states although, at the end of year 2, he conceded that it was "wholly inadequate." The site coordinator's insistence resulted in teacher distrust bordering on hostility, relieved somewhat through the intervention of the school district coordinator. The two of them ultimately developed a variant of a "good cop, bad cop" strategy that proved fairly effective. But, the project director so opposed the site coordinator's "directiveness" that the site coordinator believed in the end that he had punished her by withholding funds for her site, in effect, penalizing her site for actually managing to produce the required products.

At this site, the program's teacher-developed curriculum enhanced history courses in the participating schools, including regular history classes the Heritage Project teachers offered to students not participating in the program. Program-funded professional development and field trips introduced students and teachers to historic and cultural sites and other local resources. Among the three state sites, the strongest probability of continuation of project initiatives beyond the grant-funded period was apparent here, where the site coordinator and the school district coordinator organized a formal presentation to the local school board for that purpose.

Southwest site

In the Southwest, two participating schools within one school district were located in two culturally and historically unique communities, one predominantly Hispanic, the other entirely Native American. The Caucasian teachers at the Hispanic school, although they were voluntary participants in the program, were described by the site coordinator as inexpert and unenthusiastic, resistant to both program goals and professional development. In order to produce a curriculum document, the site coordinator, after a long struggle, resorted to writing out teachers' orally communicated ideas which, she reported, the teachers were then unwilling to implement. In contrast, the two teachers at the Native American school responded to the opportunity provided by the program with enormous energy and initiative. One teacher assumed responsibility for the program at the school when the long distance between the schools and the site coordinator created a gap in local leadership.

New principals at both schools in year 2 caused consternation. At the Hispanic school, the new principal ultimately proved supportive of the program, but he was initially viewed warily by teachers. At the Native American school, the new principal instituted sweeping school-wide changes not favorably received by the faculty. During the principal's first year, year 2 of the program, teachers filed three class action lawsuits against her. Among the plaintiffs was the quietly assertive Heritage Project lead teacher. The following year, the principal reassigned Heritage Project teachers to positions in which they could not discharge their programmatic responsibilities.

But at this site—and only this site—project personnel had forged explicit connections to the community that was the focus of the new curriculum. These teachers were supportive of and supported by the reservation community. One was a Native American, son of a tribal leader. With support of the tribe, he was reinstated to a position in which he could continue to offer the newly developed curriculum to students. But by the end of year 3, the teachers were under threat of reassignment or firing by the superintendent, to whom the principal successfully appealed. The following year, the lead teacher was reassigned off the reservation (Mabry, 1999).

The site coordinator and the teachers at the Native American school complained of the project director's lack of appreciation for local culture and context. The site coordinator also reported the project director misunderstood relative project achievement at the two schools, thinking the program at the Hispanic school stronger than that at the Native American school because of the furor at the latter, when the reverse was more accurate.

Methodology

We (Note 2) conducted an external evaluation of the program throughout years 2 and 3, 1994-96. The evaluation featured a naturalistic (Guba & Lincoln, 1989; Lincoln & Guba, 1985), responsive (Stake, 1973) approach with attention to stated project goals. Some data was collected by site coordinators, suffusing the inquiry with a participatory (Greene, 1997) combination of internal and external perspectives and improving the evaluation's potential utility (Joint Committee, 1994; Patton, 1997).

Data collection involved interviews of the project director, the central office coordinator, all site coordinators, the teachers in each participating school, and the school and district administrators. Relevant classroom activities were systematically observed and documented in all but one of the seven participating schools. (Note 3) Observation and interview data were comprehensively validated (Mabry, 1998). An extensive variety of documents was analyzed including the program proposal and interim reports to the federal funding agency, annual reports from site coordinators, curriculum documents, student products, teacher journals, assessment instruments, and materials from the project director's presentations to academic conferences about the Heritage Project. There was extensive triangulation of data by source, method, time, and observer (see Denzin, 1989). The impetus for both data collection and analysis was substantive rather than procedural in the manner advised by Erickson (1986), Lincoln and Guba (1985), Mabry (1998), Stake (1994), and Wolcott (1994), responsive to issues which emerged from the data, including:

1. How sensitive are the new social studies curricula to local history and culture?
2. How have teacher attitudes about local history and culture

influenced the focus of curriculum and instruction? Has the professionalism of teachers been enhanced?
3. To what extent has the program been invigorated by community members and institutions?
4. How well do new assessments reveal the curriculum-related achievements of students?
5. What is the extent and usefulness of networking among sites and with the project central office?

Data analysis involved review, summarization, and categorization of documents; identification of themes and refinement of issues; analytic discussions within the evaluation team and also between the evaluators and the project director and central office coordinator. Analysis featured synthesis across data types and sources, issue refinement through the constant- comparative method (Glaser & Strauss, 1967), attention to multiple perspectives, review by site personnel of descriptions and interpretations of data, and review by the project director of a draft of the report. In a final analytic thrust, the data was reconsidered according to a comprehensive theoretical framework, Bronfenbrenner's (1979) levels of ecological analysis for a system-level perspective of the ideological and practical components and relationships among components of the project.

Analysis strongly attended to data and contexts rather than to external, non-situated, general notions of program quality. Findings were emergent, thoroughly grounded in the data (Glaser & Strauss, 1967), and the data confirmed by the participants themselves as representing their experiences. That the conclusions surprised us is an indication that analysis was truly emergent, not an artifact of assumptions or bias. The thorny ethical dimensions of this study have been explored (Mabry, 1999).

A formative report was submitted to the project director for year 2 and a summative report to the federal funding agency at the end of year 3.

Discussion

This study of an initially promising, well-financed effort to restructure curriculum revealed unanticipated difficulties and limitations related to community-oriented curriculum and to teacher-generated educational reform. In a national climate of increasing attention to cultural diversity (see Banks, 1993), we expected to applaud efforts to teach children about their communities' history and to enhance their appreciation for their own cultures. We expected our results to offer empirical support for theories of educational reform which emphasize the importance of local contexts and teacher initiatives.

Instead, to the project director's extreme distress and to our personal and ideological dismay, what we reported to the funding agency and what we offer here is a presentation of intractable problems and interpretations of the nature and extent of defeating difficulties. We offer these so others may glean ideas through case-to-case generalizations (Firestone, 1993) relevant to programs of interest to them. We recognize that specific extrapolations to other cases or to the general topics of educational reform, community-oriented education, or rural education will be best made by readers who are familiar with those cases or who are experienced in those areas. We are evaluators, knowledgeable about education but not expert in each of these subfields. Our discussion of findings is organized in two ways: (1) interpretations emergent from the data related to educational reform, a vehicle for specific substantive discussion, and (2) Bronfenbrenner's (1979) levels of ecological analysis, a framework for systemic scrutiny.

Emergent interpretations

Findings from this study emphasized the importance of some difficulties regarding external support for classroom reform and for community-oriented education: (1) insensitivity of central office personnel to local conditions, (2) increased teacher workloads, (3) the cultural self-identification of curriculum developers, (4) teacher readiness and training issues, (5) a short project timeline as a condition of funding.

Insensitivity to local conditions. Centralization of project management and resources promoted coherence across sites but undermined sensitivity to local issues, with a negative net effect on achievement of project goals. The site coordinator in the Southeast, for example, complained that the project director did not (and perhaps could not) appreciate the local pace of life and educational history and traditions. She noted, for instance, that he did not take into account the lack of a teachers' union in her state and its historical result: longstanding passive resistance by teachers toward any directive from the top, including the top of the project. The lead teacher at the Native American school in the Southwest accused the project director of outright cultural insensitivity, citing among other things materials sent to the school which depicted housing and landscaping which students on the reservation were expected to understand as typical but had never experienced. In the Midwest, the relative isolation of the small towns in which the schools were located strengthened teachers' beliefs that children there needed opportunity to connect with the outside world, rather than to be focused inward as the project intended.

The project director evidenced sincerity in statements about his desire to support teachers interested in developing community-oriented curricula and in his early reports of the general agreement among participants on this fundamental goal and surprised disappointment in their lack of progress. If there was initial buy-in by teachers and site coordinators, then the Heritage Project is an example of the difficulty described by Wildavsky and Wildavsky (1984) of translating broad agreement into specific decisions involving many participants with many perspectives. The situation also bore a resemblance to the toxic discrepancy between federal expectations and local project capabilities described by DeStefano (1992).

The Heritage Project, in implementation, targeted attention more to physical artifacts of local culture, history, and traditions than to ongoing daily life. Possibly, the project director's personal interest in the sites undermined focus on the living present; site personnel murmured that he was interested in their sites for their vacation value, and the director did predict that evaluation site visits would be enjoyable. The focus might have been corrected had the project implemented the local advisory councils composed of parents and family members promised in the proposal. But, at only one site, was a half- hearted and short-lived attempt made to establish such a group. The most significant actual involvement of communities, other than a few isolated local presenters, came in the form of public displays of student work.

Ultimately in this case, the very idea of a centralized project office and its function in policy articulation and program management militated against the intimate local sensitivity implied in classroom-level reform and in community- oriented education. But funding could not have been secured without the persuasive coherence of formal aims and structures—a paradox. The dilemma in this instance raised the fundamental question of whether external or top-down support can enhance local or bottom-up reform more than illustrating how it might do so. If even well-meaning support from the top obstructs change at the classroom level, then it is unclear that current mechanisms in governmental and foundational grant funding can be counted on to support classroom-level initiatives. If centralized or top-down assistance cannot succeed at the local level, does the enormity of local diversity ensure that reform at the school and classroom level will exhibit inconsistency, a chaos of unrelated events, lack of capacity for large-scale cohesion? Even if we can accept the lack of national consensus regarding the purposes, processes, and content of education, the prospect of educational anarchy is unattractive.

Unanticipated workloads. Particularly at the Midwest site where basic project goals were resisted, teachers complained that the project's demands on their time and energy went not only beyond their expectations but also beyond their original intents and agreements. They did not feel that project-related benefits to them or to their students compensated for the additional work. Already overloaded teachers ignored the more burdensome requirements of the project, the development of customized curricula and assessments, choosing instead to continue using pre-existing social studies curricula which introduced their students to a larger conception of history and culture. They expressed frustration that the central office coordinator urged adherence to project goals and operating procedures even when these required uncomfortable expenditures of time and disregarded their deeply felt intuitions about prioritizing instructional objectives.

Further evidence of Midwest teachers' perceptions that project demands outweighed benefits came in the form of expressed desire for stipends. The intrinsic rewards of participation, when measured against the tolls exacted, were insufficient to secure their commitment to project goals. Feeling little ownership of the project, these teachers participated only nominally until pressured to do otherwise, causing frustration and resentment on all sides and lingering divisiveness.

Curriculum developers' cultural identification. At the Hispanic school in the Southwest and at the Southeast site, teachers not associated with the local groups whose history was to be emphasized in the new curricula did not exhibit the initiative to develop their own curricula. In the Southeast, the insistence of the site coordinator was sufficient to ensure the development of the required products, although she considered the curricula developed there marginally satisfactory. Caucasian teachers at the Hispanic school in the Southwest made no noticeable effort to develop or implement community-oriented curriculum.

Teacher identification with local cultural groups exerted a positive influence. At the Native American school in the Southwest, teachers were immediately energetic in their engagement with the project. The two African American teachers at the Southeast site whose history and culture were to be highlighted by new curricula displayed sensitivity and growing initiative. Working together in the Southwest site, a Native American and a Caucasian sympathizer and activist for tribal concerns exceeded project expectations by engaging in curriculum research as well as development, motivated by the opportunity to focus on the tribe's history and culture.

Teacher training. At six of the seven schools, some difficulties could be traced to teacher training. At each site, some training was provided by the project director, particularly regarding broad concepts of curriculum development and assessment, while some training was provided or arranged by the site coordinators, particularly regarding local culture and history. The latter was particularly successful in the Southeast. But there were problems both with teachers who had long been trained regarding the project goal of community- oriented social studies teaching and with teachers who had not.

The project director's former students included all three site coordinators and two of the original teachers in the Midwest. Although the project director had personally trained them and professed to have responded to their interest in community-oriented curricula, those teachers' responses to the project were the most counterproductive of all project personnel. Ironically, their earlier training with the project director had made them confident of their skills, which led to their assertiveness regarding the sufficiency of their pre-existing minor units on local history and culture and made them unresponsive to demands to emphasize local communities more. In effect, their project-relevant training predisposed them to reject the project. Professional development provided by the Heritage Project did not change their minds. (Note 4)

There were also difficulties with many of the teachers who were relatively untrained in the concepts and rationales of the project. Participating because of district directive rather than personal choice, half of the teachers in the Southeast actively resisted project goals and directives. The more receptive African American teachers had few curriculum development skills prior to the project's training and assistance but gradually and willingly improved. At the Southwest site, resistant Caucasian teachers at the Hispanic school also lacked relevant training and curriculum development skills and experience, but these deficits were overshadowed by their resistance. For unwilling teachers, professional development was marginally fruitful at best in terms of project outcomes.

The only teachers who reacted to the project with immediate enthusiasm were at the Native American school in the Southwest. Their positive efforts preceded professional development and were unaffected by it. Neither the teacher who was a member of the tribe nor the teacher who was an activist in promoting tribal issues was specifically trained in social studies teaching, but both were knowledgeable of local history and culture and energetic in pursuing project goals. This site produced the earliest and most sustained successes. The African American teachers in the Southeast, also successful in the end, posted slower results more clearly derived from professional development.

Mixed as these results were, in comparison to professional development in curriculum design, assessment training was a disaster. All personnel at all sites complained that the project director had reneged on promises to provide assessment training. The project director countered that he was merely to begin discussion of assessment, the real training to be arranged later by site coordinators. The more successful teachers managed to develop a few creditable assessments, but most teachers (and the project director) displayed little grasp of rudimentary measurement concepts. Among the major promises in the proposal to the funding agency, assessment was the worst failure.

Project timelines. In the policy hysteria (Stronach & Maclure, 1996) which has characterized educational reform since the publication of A Nation at Risk in 1983, there has been too little recognition that educational reform takes a long time, (Note 5) longer than the three years of funding granted to the Heritage Project. The project director observed that most of the first year of the project was needed for developing trust and working relationships; data suggested that a full year was not long enough for these pre-processes.

At the Southeast site and at the Hispanic school in the Southwest, community-oriented curriculum documents were created by the end of the first year, as required, only because a site coordinator wheedled or threatened teachers and, in one case, wrote the curriculum herself. At the Midwest site, curriculum documents were created by the end of the second year only because the central office coordinator demanded it, so much so that teachers complained to the project director. Most of the curriculum-based assessment instruments submitted at the end of year 3 were developed in the closing hours of the project and under duress. Few could be described as truly curriculum-based or authentic (Wiggins, 1993); most appeared unlikely to support strong inferences of student achievement as a result of the program; some were not assessments of student achievement at all, but rather instruments to evaluate instruction.

Late and marginal as many curriculum documents and assessment instruments were, it is nevertheless reasonable to ask whether too much effort was expended trying to develop them on schedule. Did personnel need the urgency of deadlines to do what they did, or would stronger curricula have been developed had they had more time? The project timeline proved to be a significant factor in ensuring that these products would be superficial and underdeveloped, artifacts providing stronger evidence of the project's failure than of its success. Here as elsewhere, the annual evaluations and interim project reports commonly expected of externally sponsored projects may actually hinder the changes funders intend to support. Premature reports document the difficulties which precede resolutions, and public exposure of growing pains demoralize program personnel. Funders' requirements for evaluations, such as ours, ensure that project personnel will be vulnerable to criticism during tender, formative periods.

An ecological analysis

Bronfenbrenner's (1979) levels of ecological analysis offer a framework for a cross-site discussion of findings in four areas: (1) macrosystem, the ideological context of the Heritage Project; (2) exosystem, the organizational and policy context; (3) mesosystem, professional interactions and relationships; and (4) microsystem, classroom interactions, practices, and relationships. Ecological analysis offers a systemic view which incorporates both beliefs and practices. From this more etic (Note 6) and abstracted perspective, the portrait of the project emanating from description at each level and from the interactions among levels highlights the conflicts which arose in the fissures between concept and implementation.

Macrosystem, the ideological context. The most significant barrier to the project's achievement of its stated goals in the Midwest was the deep but unexpected rift between project emphasis on local history and cultural traditions and teachers' beliefs about the importance of broad content for social studies instruction. Believing their students needed broader horizons, an expansive view of history, these teachers opposed a parochial approach to curriculum.

Prior to their participation in the program, the Midwest teachers had offered students a unit or two on topics such as community architecture and nearby battlefields. More concentration on local color, in their view, would have limited student learning by neglecting grand historical topics unjustifiably. This ideological conflict set the teachers against the policies and procedures emanating from the central office, adversely affected working relationships between teachers and the site coordinator, and foreclosed until the final year of the project on offering students a curriculum rich in local history and culture. Teacher resistance to externally imposed curricular change, not teacher empowerment, characterized the Heritage Project at the Midwest site for two of the three years of the project, a refusal broken during the final year only by the strong-arm tactics of the central office coordinator. Thus, conflict in the macrosystem obstructed change and also had adverse effects in all the other ecological levels. Teachers' counter-beliefs affected local policy and practice and working relationships, delaying and diminishing planned outcomes.

Exosystem, the organizational and policy context. Conflict between the central project office and every regional site seemed predictable only in hindsight. It was not initially apparent that a central structure would inhibit more than enhance classroom-level curricular reform but, in the end, the Heritage Project illustrated an inherent tension between centralized management and decentralized reform. Goals, policies, procedures, fiscal regulations, and other matters had to be formulated by the project director to exhibit the cohesion necessary to win grant funding from the federal agency and then had to be imposed on teachers in order to fulfill promises—the antithesis of teacher empowerment, local control, bottom-up strategies.

Consequently, site personnel complained of the insensitivity, even cultural insensitivity, of the project director; of being forced to do things they didn't want to do or consider appropriate; of lack of consideration of their regular teaching responsibilities; of receiving parcels of supplies so inappropriate as to be nearly unintelligible in their contexts, supplies purchased by central office personnel; of embarrassingly long delays in paying local suppliers; of broken promises regarding stipends; of their perceptions that one site was financially favored at the expense of another. These difficulties undermined trust and working relationships, with spill-over effects in terms of willingness to try the project director's ideas regarding curriculum and in terms of the nature of the delivered curriculum.

There was also conflict at the site level regarding policy and organizational practice at the Native American school in the Southwest. It appeared to the superintendent and principal that Heritage Project teachers refused to follow school and district policy, especially regarding expenditures; it appeared to the teachers and site coordinator that school and district policy and procedures were manipulated to intimidate and undermine them. It appeared to us that some simple misunderstandings might have been resolved by frank and friendly discussion, which never happened. (Note 7) These difficulties adversely affected working relationships, and teacher reassignment sharply limited student exposure to a painstakingly developed curriculum. Thus, at both the site level and the program level, conflict in the exosystem—both program policy and local policy—obstructed change and also had adverse effects in all the other ecological levels. Working relationships, confidence, and classroom practice were all undermined.

Mesosystem, professional interactions and relationships. At every regional site, site coordinators lived between one and four hours away from the schools. The project director and central office administrator were even farther away. Most participating teachers worked singly rather than as teams because their positions were in different schools and communities, no cadre or critical mass offered reinforcement or a sufficient base for secure establishment of ideas and practices. Logistically, it was difficult to bring project personnel together enough to forge strong, trusting relationships. Attempts were made: teleconferences, newsletters, site visits, student pen pals, shared videos and curriculum documents, regional professional development sessions.

Networking was not a total failure, and the second central office coordinator was particularly praised by some site personnel for her responsiveness. But every site coordinator expressed frustration with the central office. Two lamented that their difficulties with teachers were exacerbated by distance which limited site visits and support. Lack of trusting relationships inhibited adherence to project goals and directives in all but the Native American school where, analogously, lack of trusting relationships within the school district inhibited implementation of the new curriculum. Movement toward ideological and organizational harmony was as elusive as curricular change.

So, conflicts in the mesosystem at the project level and at the site level obstructed change and also had adverse affects in all the other ecological levels. Without frequent contact or strong working relationships, partnerships among site personnel, among sites, and between sites and the central office were shaky and rattled the other layers of the system.

Microsystem, classroom interactions and relationships. Fewer difficulties originated in the microsystem as the effects of difficulties in the other ecological levels ultimately struck home in classrooms. Still, at the classroom level, in about three-quarters of the schools, instructional habits, prior curricula, and entrenched pedagogies--that is, inertia--inhibited development of new curricula. Resistance to new ideas and policies put teachers at odds with project administrators. One anomaly: At the Native American school, teachers' enthusiastic development and implementation of a new curriculum initially made them the darlings of the project administrators (but not school or district administrators), admired for their ideology and for their responsiveness to project goals. Inertia was common in the microsystem, and it hindered change. Primarily, however, the microsystem, the arena in which project outcomes culminated, was adversely affected by problems in the other ecological levels.

Conclusions

As the grant period ebbed, the Heritage Project left a wash of positive effects: kids had fun and learned some worthy things about their communities which they would not otherwise have been offered; teachers got professional development and classroom materials, some of which they very much appreciated; some new working relationships were forged; some program emphases appeared sustainable at some sites. But the tide also left a disturbing debris of disappointing outcomes: primary goals perfunctorily addressed, secondary goals neglected or not attempted, curriculum and assessment products superficial or confused, feelings of frustration and resentment, and professional devaluation of an outstanding teacher. An explicit attempt to support local teachers and their desire to orient their teaching to their communities had faced and not always surmounted formidable difficulties. Embarrassed and hurt when this was reported in the final evaluation document, the project director worked determinedly to suppress and discredit the evaluation, deriding interview data—including his own—as "hearsay and innuendo."

This study shows that community-oriented education can generate teacher enthusiasm and skill development, and it can motivate student interest and learning. But our data also indicate that centralized or external support, if not carefully managed and minutely sensitive to local conditions, may poison rather than feed community-oriented educational change. Clearly, a good idea is not enough. The charm of the Heritage Project's generative notion attracted but did not sustain personnel commitment. If community-oriented curriculum was truly the desire of the teachers, a claim made by the project director but not universally confirmed by other personnel, then teachers' own ideas may not be enough to preserve the momentum for change. In the end, none of the teachers who were active in development and implementation of the new curricula were among the program conceptualizers. In contrast to the literature which favors teacher-generated or bottom-up strategies, in this case, teacher investment was unrelated to their ownership of project ideas and objectives.

Teacher commitment depended instead on a variety of factors, two of which were particularly important. One, teachers tended to be motivated by their own predictions and perceptions of benefits to students, regardless of whether the benefits devolved from an external idea. The teachers on the reservation, sure their Native American students would benefit from knowing and appreciating more about the culture and history of their tribe, engaged with enthusiasm and dedication. The teachers in the Midwest, sure their students needed to know about the wide world rather than their small towns, resisted community-oriented curricula.

Two, the cultural self-identification—or lack of it—of the curriculum developers with targeted communities proved important. African American teachers in the Southeast, whose own culture was to be celebrated in the new curricula, made more progress than their Caucasian colleagues. The Native American sympathizer and the member of the tribe performed admirably in the Southwest, but not their Caucasian counterparts in the neighboring Hispanic school. Cultural identification generated momentum and sensitivity in community-oriented curriculum development and sustainability in implementation. An important caveat: Cultural identification and investment do not automatically accompany residency. In the Midwest, where active resistance was strongest, many teachers were longstanding members of their communities.

Professional development is a politically and professionally attractive concept in the current reform climate, despite recognized limitations of professional development to effect intended change (see Little, 1994). But this study shows that professional development can bite back. The teachers most extensively trained in community-oriented education, those in the Midwest, were also most resistant to developing and implementing community-oriented education. Important as the project's ideas may have been to them, other ideas were more important. Training gave them confidence; confidence promoted autonomy; in this case, they exercised autonomy in contradiction to the program. Teacher empowerment can strengthen programs or can strengthen opposition.

This study also found professional development which was irrelevant. The reservation teachers learned what they needed on their own initiative before the project's professional development was made available; when training was offered, they found little worthwhile. The teachers most in need of training, those in the Southeast and in the Hispanic school in the Southwest, responded to it slowly or not at all. As in other human endeavors, communication proved crucial. In this case, there was a reciprocal relationship between communication and trust and problem-solving. Encumbered by logistically difficult distances, communication among far-flung partners was inadequate. Lacking frequent face-to-face contact, familiarity and trust eluded them. Also, lacking frequent face- to-face contact, problems at schools were not apprehended quickly enough by site coordinators or the project director. Small problems grow fast and unpredictably. There is a better chance of defusing irritants while they are minor, before they explode. In the Heritage Project, failure to nip unrecognized problems early undermined trust, which discouraged communication and candor in communication, which diminished the opportunity to recognize and address problems. A version of this vicious cycle was found at every site.

Our data suggest a need for much more local sensitivity and adaptability by policy-makers and educationists in colleges and universities who are interested in assisting school improvement. This study indicated that there is potential for successful initiatives originating with folks other than local implementers but also some early red flags which went unrecognized in this project: weak initial interest of some teacher-implementers, centralized rather than localized directives and decisions, logistical difficulties regarding communication and coordination, and distrust and insensitivity (often unintended). As in this case, initial levels of enthusiasm by prospective personnel may appear sufficient to justify a new initiative but may over represent their long-term commitment and fail to sustain their efforts over time.

The concepts of teacher-designed change and community-oriented education retain appeal despite the mixed results in this case. The likelihood of unintended heavy-handedness in centralized support, the difficulties of generating and maintaining personnel commitment, and the importance of enthusiasm and sensitivity emanating from teacher identification with communities suggest important considerations for these approaches. We offer this cautionary portrait to encourage rumination, discussion, and the development of increasingly sophisticated approaches to improving education in the many continuing initiatives across the country.

References

Astuto, T., Clark, D., Read, A., McGree, K., & deKoven Pelton Fernandez, L. (1994). *Roots of reform: Challenging the assumptions that control change in education.* Bloomington: Phi Delta Kappa

Educational Foundation.

Banks, J. (1993). Multicultural education: Development, dimensions and challenges. *Phi Delta Kappan*, *75* (1), 22-28.

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Darling-Hammond, L. (1990) Instructional policy into practice: "The power of the bottom over the top." *Educational Evaluation and Policy Analysis, 12* (3), 339-347.

Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

DeStefano, L. (1992). Evaluating effectiveness: A comparison of federal expectations and local capabilities for evaluation among federally funded model demonstration programs. *Educational Evaluation and Policy Analysis, 14* (2), 157-168.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York: Macmillan.

Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher, 22* (4), 16-23.

Fullan, M. (1991). *The new meaning of educational change*. New York: Teachers College Press.

Glaser, B. G. & Strauss, A. I. (1967). *The discovery of grounded theory*. Chicago, IL: Aldine.

Greene, J. G. (1997). Participatory evaluation. In L. Mabry (Ed.), *Advances in program evaluation: Evaluation and the post-modern dilemma*. Greenwich, CT: JAI Press.

Guba, E. G. & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Thousand Oaks, CA: SAGE.

Joint Committee on Standards for Educational Evaluation (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: SAGE.

Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: SAGE.

Little, J. W. (1994). Teachers' professional development in a climate of educational reform. In R. J. Anson (Ed.), *Systemic reform: Perspectives on personalizing education*. Washington, D. C.: U. S.

Department of Education Office of Educational Research and Improvement.

Mabry, L. (1998). Case study methods. In H. J. Walberg & A. J. Reynolds (Eds.), *Advances in educational productivity: Vol. 7. Evaluation research for educational productivity* (pp. 155-170). Greenwich, CT: JAI Press.

Mabry, L. (1999). Circumstantial ethics. *American Journal of Evaluation, 20* (2).

National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform*. Washington, D. C.: U. S. Government Printing Office.

Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: SAGE.

Sarason, S. B. (1990). *The predictable failure of educational reform: Can we change course before it's too late*? San Francisco: Jossey-Bass.

Seymour-Smith, C. (1986). *Dictionary of anthropology*. Boston: G. K. Hall.

Stake, R. E. (1973). Program evaluation, particularly responsive evaluation. Paper presented at conference on New Trends in Evaluation, Goteborg, Sweden. Reprinted in G. F. Madaus, M. S. Scriven & Stufflebeam, D. L. (1987), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 287-310). Boston: Kluwer-Nijhoff.

Stake, R. E. (1994). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 236-247). Thousand Oaks, CA: Sage.

Stronach, I. & Maclure, M. (1996). Mobilizing meaning, demobilizing critique? Dilemmas in the deconstruction of educational discourse. In *Cultural Studies* (vol. 1, pp. 259-276). Greenwich, CT: JAI Press.

Wiggins, G. (1993). *Assessing student performance*. San Francisco: Jossey-Bass.

Wildavsky, J. L. & Wildavsky, A. (1984). *Implementation*. Berkeley: University of California Press.

Wolcott, H. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Thousand Oaks, CA: SAGE.

Notes

1. Information about the program has been anonymized regarding title, subject area, personnel, and sites.
2. The authors were assisted with data collection by doctoral students Tracy Cronin, Jeff Davis, and Sharifah Shakirah Syed Omar.
3. In year 3, one participating teacher was attending a conference at the time of an evaluation site visit.
4. The project director disputed this, claiming that community-oriented education was initially resisted but later fully embraced by the Midwest teachers. However, seven months of classroom observations by a different researcher the following year revealed not a single creditable lesson about the local communities (anonymous, personal communication, 1998).
5. Recognition that change is slow has, of course, appeared in the literature of educational reform (see, for example, Darling-Hammond, 1990), but full understanding of the time typically needed for program development is not evident in most Requests for Proposals (RFPs).
6. An outsider's perspective based upon external categorization and structuring, as opposed to an emic approach based upon insiders' viewpoints and constructions of meaning (see Seymour-Smith, 1986, p. 92)
7. We did, however, recommend such discussion and intended to support it with our data until forbidden by the project director to do more than "watch this play out" (see Mabry, 1999).

About the Authors

Linda Mabry

Indiana University
4002 Wright Education Building
201 North Rose Avenue
Bloomington, IN 47405-1006

Telephone (812)-856.8317
fax (812)-856.8440
Email lmabry@indiana.edu

Ph.D. University of Illinois, 1995
M.S. University of Illinois, 1986
B.S. University of Houston, 1972

Linda Mabry is an assistant professor of education at Indiana University where she teaches courses in program evaluation, performance assessment, and qualitative research methodology.

Laura Ettinger

Educational Leadership & Policy Studies
Indiana University

(812) 333-8097 Email: lwoodbur@indiana.edu

Ph.D. in progress, Indiana University
M.S. Indiana University, 1993
B.A. University of Virginia, 1986

Laura Ettinger is a doctoral student in Educational Policy Studies at Indiana University where she is researching the appliation of complex adaptive systems theory to public school change. She has taught courses in the foundations of education and combines interests in education policy and school reform with program evaluation.

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs- Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
Arizona State University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa- DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

Este artículo ha sido consultado 194 veces desde el 13 de avril de 1999

Education Policy Analysis Archives

Volume 7 Number 15 Avril 13, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ
85287-0211

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Elección Educacional Entre Escuela Pública y Escuela Privada en Galicia: Un análisis comparativo del gasto de los hogares en las escuelas públicas y privadas

Prof. Dr. María Jesús Freire
Departamento de Análisis Económico
Universidad de La Coruña (España)

Prof. Dr. José Venancio Salcines
Departamento de Análisis Económico
Universidad de La Coruña (España)

Resumen
Este trabajo tiene como objeto analizar la elección educativa entre escuela pública y escuela privada para los siguientes niveles educativos: infantil, primaria y secundaria en el contexto de los hogares de Galicia (España). Este artículo también analiza en que medida está relacionado el tipo de centro educativo al que asisten los niños con la renta familiar y, la educación recibida por los padres; de esta forma, esperamos aprender algo mas acerca de la desigualdad que existe entre los dos sistemas educativos.

Abstract
The goal of the current work is to analyze educational choice between public and private schools at the pre-primary, primary and secondary levels in Galicia, Spain. We also analyze the relationship between the type of educational institution the children attend and the family income and parents' educational level. The overarching purpose of this work is to learn more about how social inequalities are mediated by public and private educational systems.

1. INTRODUCCIÓN

El análisis de la distribución de estudiantes de acuerdo al tipo de centro (público-privado) y los distintos gastos convencionales que esto conlleva para los hogares suscita actualmente un gran interés, no sólo entre los investigadores de la economía de la educación, sino también entre la clase política.

La elección educativa es un tema ampliamente discutido para tratar los problemas de la educación primaria y secundaria, pero los debates sobre la elección de mercado en la educación han estado limitados más a los aspectos teóricos que a la evidencia empírica (Levin, 1992). Las propuestas de elección educativa no implican una reducción en el papel del gobierno por lo que respecta a la financiación, pero, sin embargo, reducen su actuación en lo que respecta a la provisión (Lamdin & Mintrom, 1997). Lo que realmente es objeto de controversia es un incremento en la provisión privada que permita una mayor elección tanto para los colegios como para los padres y los estudiantes. Los argumentos a favor de la elección educativa se basan en las consecuencias favorables de la competencia. La provisión privada de educación sería un incentivo para aumentar la calidad y cantidad de educación, al disponer los estudiantes de la opción de elegir (James, 1993).

La educación, no tiene las características de un bien público (Levin, 1991), el término utilizado es de bien preferente (merit good), es decir, aquel que la sociedad estima como esencial para todos los individuos. Bajo estas circunstancias, la educación se mantiene, por lo tanto, entre dos grupos de derechos que compiten. El primer derecho, es el de los padres a elegir el tipo de colegio que desean para sus hijos. El segundo, es el derecho en una sociedad democrática de usar el sistema educativo como un medio de conseguir una mayor cohesión social.

Mientras que la intervención pública en la educación obligatoria en los niveles de primaria y secundaria es un tema que no suscita ninguna controversia, el debate reciente esta motivado por dos aspectos claves, nos referimos al sistema de financiación de los centros escolares y a la elección educativa por parte de los estudiantes o sus padres.

Las principales propuestas que se han realizado sobre los aspectos de financiación de la educación son: Milton Friedman (1962) en "Capitalism and fredom" realiza la primera propuesta sobre vouchers, aceptando la necesidad de la educación obligatoria y gratuita, el autor argumenta que la educación produce externalidades, pero esto no implica que la educación publica deba impartirse en centros públicos. Aboga por dar a los padres vouchers canjeables por una cantidad máxima de gasto por alumno en centros educativos autorizados.

La propuesta de Friedman era extremadamente simple y ha sido en general valorada por sus incentivos a mejorar la eficiencia en el sistema educativo. Sin embargo, al permitir a los padres complementar con recursos propios los vouchers ha estado sujeta a una gran controversia. Posteriormente han surgido nuevas propuestas de vouchers con algunas particularidades respecto a la inicial de Friedman.

West (1970), propone dar a los padres vouchers canjeables por una cantidad de gasto por alumno que sería idéntica, y los padres no podrían aportar financiación adicional.

Jencks (1970), plantea que los vouchers sean de mayor valor para los estudiantes de menor nivel socioeconómico. Su propuesta sirvió de base al experimento de Alum Rock en la década de los setenta.

Peacock (1983), propone unos vouchers compensatorios que estarían sujetos al impuesto sobre la renta, y tendrían menor valor para las familias de mayor renta.

El aspecto del debate actual sobre la elección educativa se basa en que el sistema educativo no produce los resultados deseados. Chubb & Moe (1990,a), señalan que la burocracia existente en los colegios públicos los hace ineficientes, esta situación se ha generado porque los directores de los colegios utilizan la mayor parte de su tiempo de trabajo en atender las necesidades burocráticas en lugar de dedicarlo a mejorar la calidad de la docencia. Los autores, señalan que, para mejorar la situación actual es necesario que se elimine la burocracia centralizada, y que los colegios dispongan de más autonomía. Su trabajo, tanto en lo que se refiere a los supuestos teóricos como al análisis empírico ha recibido muchas críticas (Henig, 1994). Numeroso autores manifiestan que el problema de la educación pública americana ha sido exagerado (Berliner & Biddle, 1995) y muestran serias dudas sobre la información disponible, respecto al incremento en el gasto público en educación.

En el análisis sobre el debate de la elección educativa es necesario distinguir dos tipos de evidencia: la directa y la indirecta.

Al hablar de la evidencia directa se hace referencia a programas de elección educativa que se han puesto en marcha en determinados distritos educativos o entre ciertos colectivos de estudiantes, nos referimos a los experimentos con vouchers que se llevaron a cabo en Estados Unidos de América en los años setenta en el distrito de Alum Rock (California), y desde 1990 en Milwaukee (Wisconsin), y en el Reino Unido desde 1981, donde opera el programa "Assisted Places Scheme". Los experimentos con vouchers tratan de introducir la competencia en el sistema educativo entre los centros públicos y privados, y acabar con la asignación automática de los estudiantes a las escuelas de acuerdo con su zona de residencia.

En la literatura sobre la elección educativa, la evidencia indirecta compara la escuela pública y privada y presenta un abanico muy amplio de temas de discusión. Estos análisis son de interés por muchas razones. Algunos trabajos analizan los resultados obtenidos en las escuelas privadas respecto a las públicas. Otros trabajos, examinan los efectos que generaría la competencia en el sector educativo. Por último, también se estudia la elección educativa entre escuela pública y privada teniendo en cuenta una serie de variables relevantes.

El objetivo de nuestro artículo es determinar y clasificar las diferencias en los gastos educacionales de los hogares gallegos para los siguientes niveles educativos: infantil, primaria y secundaria considerando dos tipos de escuelas (pública y privada) del sistema educativo existente en España. Este artículo también incorpora como variables independientes la renta familiar disponible; los niveles educativos de los padres y el origen familiar. Estas variables se consideran fundamentales para estudiar la distribución de alumnos por tipo de centro, sin embargo, también podrían ser analizadas otras variables relevantes como la habilidad. La fuente utilizada para la obtención de los datos ha sido la Encuesta de Presupuestos Familiares 90/91 (EPF) (Nota 1) realizada por el Instituto Nacional de Estadística de España. Esta encuesta recoge los gastos y las características mas relevantes de 21.155 hogares españoles que para el objetivo de nuestro trabajo en Galicia (España) quedan reducidos a 1.763. Los datos sobre el número de alumnos en los distintos niveles educativos han sido facilitados por la "Oficina de Educación del Gobierno de Galicia", y corresponden al curso académico 1995/96. El gasto anual por estudiante ha sido tipificado en gastos convencionales y gastos comunes. Los niveles educativos considerados son: infantil (de 3 a 5 años), primaria (de 6 a 11 años) y secundaria (de 12 a 13 años) (Nota 2). Los gastos convencionales comprenden: a) Matrícula y pagos por inscripción, b) Pagos en concepto de mensualidad. Los gastos comunes comprenden: a) Pagos complementarios o extra-escolares, b) Gastos en libros y fichas, c) Gastos en material escolar, d) Seguro escolar y revisiones realizadas exclusivamente en los centros de enseñanza, e) Comedor escolar, f) Asociación de padres de alumnos, y g) Transporte escolar.

2. LAS ECONOMÍAS DEL BIENESTAR DE LA EDUCACIÓN

La intervención pública en el sector educativo para financiar una educación básica y obligatoria, apenas es discutible. Sin embargo, el tipo de intervención suscita un gran debate político.

Los atributos de externalidades de la educación son utilizados frecuentemente para defender la intervención estatal en la educación, ya que los beneficios privados que obtiene el estudiante son menores que los beneficios sociales.

Por otra parte, el problema de información imperfecta, por parte de los demandantes, conducen a la obtención de resultados subóptimos en los mercados privados. Los demandantes (estudiantes y padres) no aprecian el valor de la educación para el futuro y la cantidad demandada sería menor que con información perfecta, bajo esta perspectiva, la actuación del sector público, trata de incrementar la demanda educativa. Pero, la provisión pública puede conducir a unos gastos educativos excesivos, ya que existen otros métodos de provisión más eficientes.

Los fallos de mercado, mencionados anteriormente, no conducen de forma racional a una intervención del sector público en la provisión y financiación de la educación, sin embargo, el principio de equidad justifica la actuación del gobierno en lo que respecta a la financiación de la educación básica. Sin financiación pública, los estudiantes procedentes de familias de rentas bajas no podrían acceder a los beneficios que genera la educación, y la política educativa al mejorar la equidad podría ser un instrumento muy eficaz de redistribución de renta (Glomm & Ravikumar, 1992). Pero, incluso teniendo en cuenta el principio de equidad, la intervención del sector público se ceñiría a la financiación de la educación y no a la provisión.

Las razones que avalan la intervención pública en el sector educativo no generan acuerdo en la literatura económica. Para justificar la educación obligatoria y gratuita (en España, a partir de la LOGSE, 1990, se extiende hasta los 16 años) algunos autores (Friedman, 1962) destacan los efectos externos que la educación de un individuo tiene sobre la sociedad democrática y estable. Otros autores

(Stiglitz, 1988) argumentan que la educación genera importantes beneficios privados, pero la intervención pública también esta justificada por objetivos distributivos, y asegura una distribución de oportunidades educativas socialmente deseable.

El hecho de que existan imperfecciones en el mercado de capitales que no permitan a los jóvenes financiar las inversiones en capital humano tiene consecuencias tanto para la eficiencia como para la distribución. Por otra parte, la falta de información por parte de los estudiantes y las familias les podrían conducir a adquirir menos educación de la que en el mercado de trabajo, en un futuro, van a necesitar.

Serán por tanto los principios de equidad y eficiencia los que marquen las actuaciones de política educativa. Estos han sido estudiados habitualmente como principios en conflicto. Se dice que un incremento en la equidad lleva a una pérdida de eficiencia e impone costes superiores (Nota 3). El medio para poder clasificar la posible configuración de la educación se apoya en la cuantificación de los gasto de los hogares gallegos, tanto en la escuela pública como en la escuela privada.

Dado el objetivo de nuestro trabajo, el primer paso consiste en conocer la distribución de estudiantes por tipo de centro, ya que esto nos permitirá obtener la demanda agregada de plazas escolares. La situación actual en Galicia, en los niveles educativos considerados y por tipo de centro es la siguiente: infantil (61.729 alumnos), primaria (173.468 alumnos) y secundaria (78.023 alumnos). La población en edad escolar referidas a estos niveles educativos, asciende a 313.220 alumnos. Un análisis pormenorizado de la distribución de alumnos por tipo de centro se puede ver en la tabla 1, y nos muestra que alrededor del 70% de los estudiantes realizan sus estudios en centros públicos.

Tabla 1.Distribución (%) de estudiantes por tipos de escuela en Galicia

Nivel Educativo	Pública	Privada
Infantil	72%	28%
Primaria	71	29
Secundaria	72	28
Media	71	29

Estos datos que muestran la relevancia de los centros públicos, están justificados por la diseminación de la población, y el importante peso del entorno rural. En este trabajo intentaremos cuantificar el gasto anual por estudiante de acuerdo al tipo de escuela (pública o privada) a la que asisten los jovenes.

Tabla 2. Gasto anual por estudiante de los hogares gallegos en escuela privada por nivel educativo (US dollars).

Nivel Educativo	Nº de Estudiantes	Gastos Convencionales	Gastos Comunes	Gasto Total
Infantil	17.370	699	109	808
Primaria	50.613	425	205	629
Secundaria	22.132	495	157	652
Total	90.115	495	174	669

Tabla 3. Gasto anual por estudiante de los hogares gallegos en escuela pública por nivel educativo (US dollars).

Nivel Educativo	Nº de Estudiantes	Gastos Convencionales	Gastos Comunes	Gasto Total
Infantil	44.359	30	110	140
Primaria	122.855	0	205	205
Secundaria	55.891	0	157	157
Total	223.105	6	174	180

Las tablas 2 y 3 resumen el gasto anual por estudiante para los hogares de Galicia en las escuelas pública y privada. Los datos muestran que los gastos de enseñanza para los hogares con miembros en las escuelas privadas son mas altos que en las escuelas publicas (el mas alto corresponde a educación infantil (US$ 808) (Nota 4), seguido de los gastos en educación secundaria (US$ 652) y en primaria (US$ 629)).

El siguiente paso consiste en analizar la hipótesis de que los hogares gallegos tuviesen que financiar todos los gastos relacionados con la educación de sus hijos, y la incidencia que esta situación hipotética tendría sobre los ingresos de los hogares. Ante esta hipótesis la cuestión no es tanto que tipo de escuela eligen las familias (pública o privada) sino la que pueden elegir teniendo en cuenta sus ingresos disponibles.

Para ello, es necesario contrastar el gasto anual por estudiante con los datos de renta (Nota 5) de los hogares gallegos por decilas de ingreso o renta (Nota 6). En el siguiente paso analizamos la reducción en los ingresos disponibles de los hogares (clasificados por decilas de ingreso).

La tabla 4 (Nota 7) muestra como la introducción de un nuevo sistema de financiación de la educación en los niveles de escuela infantil, escuela primaria y escuela secundaria, donde los hogares gallegos soportasen el gasto anual por estudiante tendría un fuerte carácter regresivo, penalizando a los hogares con las rentas mas bajas, especialmente al 20% de los hogares de menos ingresos.

Tabla 4. Incidencia del gasto anual, por estudiante y nivel educativo, sobre la renta de los hogares gallegos (US dollars)

Decilas renta	Renta (US dollars)	Incidencia (%) infantil	Incidencia (%) primaria	Incidencia (%) secundaria
1ª decila	5.264	13	8	9
2ª decila	9.917	7	4	5
3ª decila	13.469	5	3	4
4ª decila	16.865	4	3	3
5ª decila	20.128	3	2	2
6ª decila	24.118	3	2	2
7ª decila	29.088	2	1	2
8ª decila	35.305	2	1	1
9ª decila	44.751	1	1	1
10ª decila	69.891	1	1	1

Este carácter regresivo es mas pronunciado en aquellos hogares que se encuentran en la primera decila, donde los ingresos familiares caerían un 13%. En los hogares que se encuentren en la segunda decila se observa que la reducción en el ingreso alcanzaría un 7%, cuando nos referimos a la financiación de la educación infantil. La caída en los ingresos en las siguientes decilas sería menor del 5% en todos los demás casos.

Si no existiesen transferencias públicas, una cantidad de educación menor que la óptima sería adquirida por algunos hogares , (Nota 8) y la sociedad como un todo sufriría las consecuencias. Como ha quedado expuesto anteriormente, la provisión de educación pública, reduce el coste privado para muchas familias y aumenta el tipo de rendimiento marginal privado hasta igualarlo con el tipo de rendimiento marginal social. No obstante, la evidencia muestra que el argumento de los beneficios externos para la financiación pública de la educación en estos niveles educativos puede ser entendido como una mejora de la ciudadanía. La justificación utilizada con mayor frecuencia para la financiación pública de la educación es que muchos hogares no serían capaces de soportar con sus ingresos el gasto educacional de sus hijos y tampoco lo considerarían una prioridad , (Nota 9) la educación pública viene a llenar este vacío.

3. RELACIÓN ENTRE LAS CARACTERÍSTICAS DEL PADRE Y EL NIVEL EDUCATIVO DE LOS JÓVENES

Se analizan las características socioeconómicas y socioprofesionales y el nivel de estudios alcanzado por el sustentador principal (padre), (Nota 10) y la distribución de estudiantes con respecto a los niveles educativos considerados en este trabajo. Para este propósito se adopta como punto de partida la teoría del capital humano, que concibe la educación como una forma específica de capital. El hecho de que este tipo de capital sea inseparable de la persona, explica que se le denomine capital humano. Así, la demanda de educación constituye una demanda de inversión cuya característica básica es que incrementa la producción potencial de los individuos.

La demanda de educación se acrecienta de generación en generación con el cambio en las aptitudes de la sociedad y con la convicción generalizada de que la educación constituye un vía segura para el acceso a mejores oportunidades futuras tanto monetarias como no monetarias. Los deseos de las familias de adquirir mayor educación para sus hijos, ha acrecentado las ayudas estatales en todos los niveles educativos. Esto ha permitido superar las restricciones presupuestarias a todos los segmentos sociales, convirtiendo los deseos en demanda efectiva de plazas educativas.

El análisis de la condición socioprofesional (SC) (Nota 11) del padre se ha realizado utilizando la clasificación metodológica de la EPF . (Nota 12) La observación de la distribución de alumnos en los diferentes niveles educativos teniendo en cuenta la SC (ver tabla 5), muestra que la mayoría de los alumnos (educación infantil, primaria y secundaria) pertenecen a hogares donde el padre desempeña su profesión como trabajador por cuenta ajena en los sectores secundario y terciario (categorías 1 y 2). La distinción dentro de la SC entre trabajadores manuales y no manuales nos muestra que los estudiantes se distribuyen por igual entre estas dos categorías. Esta afirmación no es extensiva al nivel educativo de secundaria, donde los sustentadores principales (padres), pertenecientes al grupo de trabajadores manuales en los sectores de la industria y servicios (30.77%) es superior al grupo de trabajadores no manuales en estos mismos sectores (22.76%).

Tabla 5. Distribución (%) de estudiantes por niveles educativos teniendo en cuenta la condición socioprofesional de el padre (SC)

SC	Infantil	Primaria	Secundaria
1	33.62	29.43	30.77
2	31.90	29.69	22.76
3	18.97	17.19	21.15
4	4.31	7.81	7.69
5	0.86	1.82	4.17
6	10.34	14.06	13.46
Total	100.00	100.00	100.00

Los datos obtenidos para las categorías socioprofesionales de la EPF son similares a los datos del mercado de trabajo publicados en la encuesta de población activa (Nota 13) para Galicia (EPA). Sin embargo, la mayor aproximación se observa en la distribución sectorial de los trabajadores en las provincias gallegas con menor peso en la agricultura.

En los datos de la tabla 5 es necesario resaltar el escaso peso de alumnos en educación infantil para la categoría 4—trabajadores de la agricultura—además, los padres dedicados a la agricultura tienen el nivel de estudios mas bajo. Esta situación provoca la tardía incorporación de los niños del entorno rural a la vida escolar, ya bien por el aislamiento geográfico o por la distancia a la escuela.

Tabla 6. Distribución (%) de estudiantes por niveles educativos teniendo en cuenta la condición socioeconómica del padre (SEC)

SEC	Infantil	Primaria	Secundaria
1	0.88	4.93	4.48
2	6.14	4.38	4.14
3	19.30	18.90	23.79
4	14.04	10.68	4.83
5	20.18	26.30	23.45
6	36.84	32.88	36.55
7	0.88	0.82	2.07
8	1.75	1.10	0.69
Total	100.00	100.00	100.00

La condición socioeconómica (SEC) (Nota 14) del padre también debe ser estudiada . (Nota 15) Al analizar la distribución de los alumnos en los distintos niveles educativos teniendo en cuenta el SEC, (ver tabla 6), se observa la aglutinación de los estudiantes en cuatro categorías. En los tres niveles educativos destacan los hogares donde el padre tiene la categoría de contramaestre, operario cualificado o miembro de cooperativas no agrarias. Esta situación se justifica porque de todas las categorías socioeconómicas analizadas, (ver tabla 7), la categoría 6, mencionada anteriormente, es la que aglutina el mayor número de hijos estudiando (72.64%). A esto habría que añadir que en esta categoría se concentra el mayor número de hogares (29.53%). La siguiente categoría en importancia es el resto de personal de administración y servicios y profesionales de las fuerzas armadas.

Tabla 7.Hogares con hijos estudiando, acorde con la condición socioeconómica del padre (SEC)

SEC	1	2	3	4	5	6	7	8	Media
% de hijos	39.2	51.2	70.4	57.4	66.7	72.6	63.6	47.4	64.4

Un dato que a priori podría llamar la atención es la pérdida paulatina de importancia de aquellos estudiantes pertenecientes a los hogares donde la categoría socioprofesional del padre es la 4ª, que se corresponde con directores profesionales o jefes administrativos por cuenta ajena.

Otra característica que, en términos de la teoría del capital humano, puede considerarse como relevante es el nivel de estudios alcanzado por los padres. Parece evidente que el nivel educativo de los padres, a través de la transmisión de conocimientos a sus hijos y la existencia de un mayor o menor ambiente cultural en el hogar, constituye un factor importante para la futura formación de sus hijos. En esta sección como en las anteriores, se ha utilizado la codificación de la EPF . (Nota 16) Cuando tratamos de interpretar los resultados de la distribución de estudiantes de acuerdo con el nivel de estudios del padre , (Nota 17) (ver tabla 8), se puede observar que los padres con nivel de estudios primarios y secundarios constituyen los grupos con el mayor número de estudiantes; esta situación se mantiene para todos los niveles educativos analizados. El rasgo característico de los recursos humanos de Galicia, donde el nivel de estudios de la población adulta se concentra en estudios primarios y secundarios, es similar a lo que ocurre al nivel de España. (Nota 18)

La evolución de los años de escolarización de la población española, en las tres últimas décadas, ha tenido una espectacular mejora educativa, casi se han duplicado los años medios de escolarización de la población adulta. En nuestro trabajo, tabla 8, se observa como los padres de más edad que no han disfrutado de los beneficios de la educación gratuita, y tienen a sus hijos cursando el nivel de educación secundaria, el 61,2% de esta población adulta tiene como máximo nivel de estudios primarios. Sin embargo, los padres más jóvenes, que tienen a sus hijos en el nivel de estudios de educación infantil, tan sólo el 33,8% tienen como máximo nivel de estudios primarios (ver tabla 8).

Tabla 8. Distribución de estudiantes (%) acorde al nivel de estudios del padre

Nivel de estudios del padre	0	1	2	3	4	5	6	7	8	9	Total
Infantil	0.9	2.6	30.3	26.7	9.5	8.6	5.1	6.9	5.2	4.3	100
Primaria	0.8	6.8	41.4	19.8	9.6	2.9	4.9	2.3	7.3	4.1	100
Secundaria	1.3	12.8	47.1	21.5	3.8	1.0	2.5	4.1	4.5	1.3	100

En la tabla 9, presentamos el gasto de los hogares gallegos en educación infantil condicionado al nivel de estudios del padre y la distribución de alumnos en centros públicos y privados. Un primer aspecto a destacar es, la diferencia entre el nivel de estudios conseguido por el padre y el tipo de centro al que acuden los hijos. A medida que aumenta el nivel de estudios del padre se incrementa la demanda de educación para los hijos en centros privados. Al enfocar las características especificas de cada grupo, cabe señalar que los padres con nivel de estudios 1 y 2 tienen porcentajes similares dentro de la distribución de alumnos por tipo de centro. Es decir, los padres sin estudios y con estudios de primaria mantienen el mismo peso relativo sin grandes diferencias entre centros públicos y privados. Los padres que han alcanzado el nivel de estudios de primaria que representan el 38,3% de los padres con hijos estudiando en Galicia. Este grupo de padres son los más representativos en la distribución de alumnos por tipo de centro, y esta situación, lógicamente, se traslada al cuadro de gasto de los hogares en educación, donde representan el grupo con mayor gasto realizado.

Tabla 9. Gasto de los hogares gallegos en educación infantil condicionada al nivel de estudios del padre (US dollars)

	Distribución de estudiantes (%)		Gasto de los hogares (US dollars)	
Nivel de estudios	Pública	Privada	Pública	Privada
0	1	0	76.466	0
1	2	3	152.932	400.975
2	30	31	1.835.179	4.410.722
3	31	17	1.911.645	2.405.848
4	9	11	535.260	1.603.899
5	11	3	688.192	400.975
6	5	6	305.864	801.949
7	5	11	305.864	1.603.899
8	5	6	305.864	801.949
9	1	11	764.666	1.603.899
Total	100	100	6.193.728	14.034.115

En la tabla 10, analizamos el gasto de los hogares gallegos en educación primaria. Los datos obtenidos se presentan muy similares a los observados en educación infantil. No obstante, existen características propias que merecen ser comentadas. En primer lugar, en la distribución de alumnos en las escuelas privadas son poco representativos los padres con nivel de estudios primarios, en segundo lugar, hay un incremento en los centros privados del porcentaje de alumnos cuyos padres han completado los estudios de bachiller y de nivel medio universitario (carrera universitaria de tres años: diplomatura, ingeniería técnica, arquitectura técnica). También se observa, un incremento del porcentaje de alumnos con padres que tienen nivel de estudios universitarios de grado superior (carrera universitaria de cinco años: licenciatura, ingeniería superior, arquitectura superior).

Tabla 10. Gasto de los hogares gallegos en educación primaria condicionada al nivel de estudios del padre (US dollars)

	Distribución de estudiantes (%)		Gasto de los hogares (US dollars)	
Nivel de estudios	**Pública**	**Privada**	**Pública**	**Privada**
0	1	0	264.404	0
1	9	1	2.203.372	321.763
2	47	26	11.721.939	8.365.856
3	21	16	5.288.093	5.148.219
4	7	16	1.850.832	5.148.219
5	2	4	616.944	1.287.054
6	5	5	1.233.888	1.608.818
7	2	3	528.809	965.291
8	5	14	1.233.888	4.504.691
9	1	14	176.266	4.504.691
Total	100	100	25.118.441	31.854.607

Tabla 11. Gasto de los hogares gallegos en educación secundaria condicionada al nivel de estudios del padre (US dollars)

	Distribución de estudiantes (%)		Gasto de los hogares (US dollars)	
Nivel de estudios	**Pública**	**Privada**	**Pública**	**Privada**
0	2	0	143.403	0
1	15	4	1.326.482	636.283
2	52	31	4.517.209	4.453.985
3	19	31	1.649.140	4.453.985
4	2	10	179.254	1.484.661
5	1	0	107.552	0
6	2	4	179.254	636.283
7	2	10	215.105	1.484.661
8	4	6	358.508	848.378
9	1	3	71.701	424.189
Total	100	100	8.747.612	14.422.429

Cuando nos remitimos al gasto de los hogares gallegos en educación secundaria, (ver tabla 11), entramos obviamente, en una situación donde el padre o sustentador tiene una edad media más alta; esto implica un mayor peso de padres con nivel de estudios 1,2 y 3. Esta situación ocurre por igual en ambos tipos de escuelas (públicas y privadas). No obstante, los estudiantes que acuden a centros privados tienden a vivir en hogares donde el padre ha alcanzado un mayor nivel de estudios.

4. EFICIENCIA VERSUS EQUIDAD

Los economistas de la educación, Blaug (1985), ya no creen que el concepto de demanda privada de educación proporcione una base suficientemente sólida como para permitir una planificación cuantitativa (dentro del sistema público) y, se plantean volver a examinar los modelos de financiación dominantes; incluso el análisis de las tasas de rendimiento no se consideran fundamentales para la elaboración de políticas públicas. La opción más defendida es que, la educación alcanzada afecta tanto a la distribución de la renta como al nivel académico alcanzado por la población y que según como se financie la educación puede aumentar las desigualdades observadas en los ingresos de los individuos.

La relación positiva entre mayores niveles de renta y niveles de educación, constituye uno de los descubrimientos más sorprendentes de la ciencia social moderna . (Nota 19) Sin embargo, se observa una excepción en la educación profesional, Psacharopoulos (1980). Este nivel educativo, formación profesional, genera, según se demuestra en la investigación anteriormente citada, menores tasas de rendimiento monetario para los estudiantes y ofrece una probabilidad menor de encontrar empleo, y una mayor probabilidad de que sus ingresos sean menos elevados.

El carácter regresivo que conllevaría para los hogares gallegos financiar con recursos propios la educación de sus miembros, vulnera por si mismo el principio de equidad, y también anula el carácter redistributivo que sobre la renta se observa en la trayectoria de ingresos de ciclo vital, Conlisk (1971).

Analizando la tabla de contingencia del nivel de estudios del padre por decilas de renta, (ver tabla 12), se obtiene la relación entre las variables nivel de estudios del padre y la renta por decilas. Los datos obtenidos permiten constatar que el 57,7% de los hogares gallegos que tienen el cabeza de familia analfabeto (nivel de estudios=0) se sitúan en la primera decila de renta; de hecho, la inmensa mayoría de los padres analfabetos, 73,1%, están ubicados en las tres decilas más bajas de renta. Esta relación causal entre nivel de estudios del padre y nivel de renta no es sorprendente para el investigador. Ha sido teórica y empíricamente demostrada en la literatura de la economía de la educación, la existencia de esta relación, pero las diferencias culturales, económicas y sociales de los distintos países siguen exigiendo la realización de contrastaciones.

Si analizamos la situación de los padres con estudios universitarios en Galicia observamos una situación opuesta a la anteriormente descrita. La inmensa mayoría de estos padres, 65,6%, están agrupados en las tres últimas decilas de renta. La senda que se recorre desde el nivel de estudios=0 hasta nivel de estudios=9 es coherente y esta en consonancia con los desarrollos teóricos existentes en la economía de la educación. En la tabla 12, se puede observar los incrementos paralelos entre el nivel de estudios alcanzado por el padre y el nivel de renta.

Tabla 12. Distribución (%) del nivel de estudios por decilas de renta.

Fila % Colum. %										
Decila	0	1	2	3	4	5	6	7	8	9
1	17,2	51,1	23,6	2,9	0,6	1,1	0,6	1,1	1,1	0,6
	57,7	19,2	6,2	2,4	1,2	5,4	2,4	5,3	2,3	1,6
2	1,7	40,8	42,0	5,7	2,9	1,7	1,7	0,0	2,3	1,1
	5,8	15,3	11,0	4,8	6,0	8,1	7,3	0,0	4,5	3,3
3	2,9	33,9	41,4	10,3	2,9	0,0	2,3	1,7	3,4	1,1
	9,6	12,7	10,8	8,6	6,0	0,0	9,8	7,9	6,8	3,3
4	1,7	28,2	37,9	19,5	4,6	1,7	1,1	1,7	2,9	0,6
	5,8	10,6	9,9	16,3	9,6	8,1	4,9	7,9	5,7	1,6
5	0,6	25,9	44,3	10,9	5,2	5,2	1,1	4,6	1,1	1,1
	1,9	9,7	11,6	9,1	10,8	24,3	4,9	21,1	2,3	3,3
6	0,6	20,1	45,4	12,6	8,6	1,1	2,3	1,7	4,6	2,9
	1,9	7,5	11,9	10,5	18,1	5,4	9,8	7,9	9,1	8,2
7	0,6	17,8	44,3	16,1	3,4	4,6	2,9	2,3	3,4	4,6
	1,9	6,7	11,6	13,4	7,2	21,6	12,2	10,5	6,8	13,1
8	2,3	18,4	40,2	17,2	4,6	1,7	2,9	2,9	4,0	5,7
	7,7	6,9	10,5	14,4	9,6	8,1	12,2	13,2	8,0	16,4
9	1,7	16,7	35,6	12,1	8,0	1,1	4,0	4,6	9,2	6,9
	5,8	6,3	9,3	10,0	16,9	5,4	17,1	21,1	18,2	19,7
10	0,6	13,9	28,3	12,7	6,9	2,9	4,6	1,2	18,5	10,4
	1,9	5,2	7,4	10,5	14,5	13,5	19,5	5,3	36,4	29,5

5. UN ANALISIS DE COMPONENTES PRINCIPALES

La matriz de información hace referencia al gasto de los hogares gallegos en educación, recogidos en la Encuesta de Presupuestos Familiares. Se han seleccionado los hogares con hijos estudiando en los niveles educativos analizados en este trabajo. Las variables analizadas son las descritas a continuación:

- Preesct: Gastos totales en educación infantil por hogar.
- Ged: Gasto acumulado del hogar en educación.

- Egbimt: Gasto total del hogar en educación primaria.
- Egbst: Gastos totales del hogar en secundaria.
- Apas: Gastos totales del hogar en asociaciones de padres.
- Comedor: Gastos totales del hogar en comedor.
- Egbim2: Gastos totales del hogar en centros privados de educación primaria.
- Egbs2: Gastos totales del hogar en centros privados de educación secundaria.
- Trin: Gastos totales del hogar en transporte interurbano.
- Trur: Gastos totales del hogar en transporte urbano.
- Preesc1: Gastos totales del hogar en centros públicos de educación infantil.
- Preesc2: Gastos totales del hogar en centros de educación privados de educación infantil.

Tabla 13. Initial Statistics

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
PREESCT	1,00000	*	1	3,62311	30,2	30,2
GED	1,00000	*	2	1,86731	15,6	45,8
EGBIMT	1,00000	*	3	1,61962	13,5	59,3
EGBST	1,00000	*	4	1,01716	8,5	67,7
APAS	1,00000	*	5	1,00248	8,4	76,1
COMEDOR	1,00000	*	6	,93199	7,8	83,8
EGBIM2	1,00000	*	7	,67845	5,7	89,5
EGBS2	1,00000	*	8	,65227	5,4	94,9
TRIN	1,00000	*	9	,28851	2,4	97,3
TRUR	1,00000	*	10	,19057	1,6	98,9
PREESC1	1,00000	*	11	,09020	,8	99,7
PREESC2	1,00000	*	12	,03833	,3	100,0

La tabla anterior muestra las comunalidades iniciales y autovalores de la matriz de correlaciones. Dado que la variabilidad total del fichero de datos puede explicarse a partir de los 12 componentes expuestos en el cuadro anterior, tenemos como consecuencia que las comunalidades iniciales son iguales a 1 para todas las variables objeto de estudio. Se puede observar igualmente en el cuadro anterior que tenemos 5 autovalores con valores superiores a la unidad. Este dato será de importancia posterior, ya que únicamente hemos considerado componentes principales a aquellos factores con un autovalor superior a la unidad.

El primer factor de la tabla anterior nos muestra un valor de

3,62311, y el sólo es capaz de explicar el 30,2% de la variabilidad del fichero. Los dos siguientes factores presentan valores parejos, 1,86 y 1,62, y son capaces de explicar respectivamente el 15,6% y el 13,55% de la variabilidad total del sistema. El cuarto y el quinto factor presentan igualmente valores superiores a la unidad, 1,0716 y el 1,00248, y pueden explicar el 8,5% y el 8,4% restante de la variabilidad del sistema.

Estos cinco factores explican el 76,1% de la variabilidad de la muestra definida anteriormente. El resto de factores, desde el sexto hasta el duodécimo, presentan valores inferiores a la unidad y, por tanto, no serán considerados componentes principales.

Tabla 14. Factor Matrix

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
PREESCT	.68800	-.28179	.64429	-.00459	.00334
GED	.89550	.22038	.07339	.02365	.03742
EGBIMT	.78229	.02030	-.52436	.04567	.06258
EGBST	.06237	.91099	.22890	.00673	.01398
APAS	.21369	.13235	-.21355	-.34383	-.02389
COMEDOR	.67522	-.04170	-.22802	.13838	.03743
EGBIM2	.68231	.03643	-.45121	-.04512	.02482
EGBS2	.03022	.89055	.23516	-.07085	.02989
TRIN	-.00185	.10645	-.01363	.92958	.03643
TRUR	.63623	.05424	-.05273	-.05551	-.05166
PREESC1	.00574	-.09026	.14545	-.05272	.98165
PREESC2	.59925	-.27010	.70084	-.00646	-.16079

Estas puntuaciones factoriales también pueden ser consideradas como los coeficientes de correlación lineal de Pearson entre el factor y la variable implicada. Por tanto, consideraremos las puntuaciones factoriales como coeficientes de correlación entre variables y factores.

La primera variable muestra una mayor correlación con los factores 1 y 3, caso similar se da para el gasto de los hogares gallegos en educación infantil impartida en centros privados. Este hecho esta relacionado con el importante peso que tiene la variable preesc2 dentro de preesct. Esta correlación entre los factores 1 y 3 no se presenta en la variable preesc1, que muestra la puntuación factorial más alta de toda la matriz factorial con la componente principal que menor porcentaje de variabilidad explica, el factor 5.

La variable gasto de los hogares en educación primaria tanto si se hace referencia a los totales por hogar como a los correspondientes a centros privados presentan un importante grado de correlación con el factor 1. Las variables del factor 1, comedor y trur, obtienen valores relevantes. No obstante, las mayores puntuaciones factoriales observables en toda la matriz factorial se dan para las variables egbs2 y el factor 2 (puntuación factorial=0,89055), egbst y el factor 2 (puntuación factorial= 0,91099) y la variable preesc1 y el factor 5 (puntuación factorial=0,98615). En la siguiente tabla se muestran las comunalidades finales, explicadas por las componentes seleccionadas.

Tabla 15. Final Statistics

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
PREESCT	.96789	*	1	3.62311	30.2	30.2
GED	.85783	*	2	1.86731	15.6	45.8
EGBIMT	.89335	*	3	.61962	13.5	59.3
EGBST	.88642	*	4	1.01716	8.5	67.7
APAS	.22758	*	5	1.00248	8.4	76.1
COMEDOR	.53020	*				
EGBIM2	.67311	*				
EGBS2	.85520	*				
TRIN	.87697	*				
TRUR	.41626	*				
PREESC1	.99575	*				
PREESC2	.94912	*				

Estas estadísticas explican la mayor proporción de variabilidad en las siguientes variables: 0,96789 en la variable preesct; 0,99575 en preesc1; 0,94912 en la variable preesc2; 0,899335 en egbimt: 0,87697 en trin; 0,85783 en ged y 0,85520 en egbs2.

6. CONCLUSIONES

La elección educativa (escuela pública versus escuela privada) en España es un tema de gran interés tanto para la clase política como para los investigadores de la economía de la educación desde hace dos décadas. Sin embargo, los estudiosos del tema consideran necesario continuar investigando con nuevos programas alternativos de elección educativa (un sistema combinado de centros públicos y privados).

En teoría el sistema de vouchers es el que daría mayores posibilidades de elección a padres y estudiantes. Permitiría mejorar la calidad de la educación al introducir competencia en el sector.

El sistema de elección educativa existente podría decirse que es perverso, debido, entre otros motivos, a la importancia que se concede a la zona de residencia familiar, y a la escasa información disponible sobre la calidad de los centros educativos.

El sistema de financiación de la enseñanza obligatoria, aprobado en España desde 1985 trata de garantizar el derecho de los padres de poder optar entre los centros que reciben financiación pública.

En este trabajo se analizó el gasto anual por estudiante para ambos tipos de centro, y los resultados obtenidos permiten concluir que las escuelas privadas son más caras para las familias que los centros públicos. Si todos los hogares gallegos tuvieran que soportar el gasto anual de la educación de sus hijos, se podría llegar a expulsar del sistema educativo a aquellos alumnos pertenecientes a los hogares de rentas más bajas.

Otra conclusión que se puede extraer de este análisis es la importancia/necesidad de mantener el actual sistema de financiación público, pero impulsando la investigación para disponer de más referencias respecto a la provisión privada, con el fin de alcanzar una mayor competencia que conduzca a un marco de menores costes y mayor elección. Un sistema educativo compuesto exclusivamente por centro privados, donde las familias tuviesen que financiar todos los gastos educativos, perjudicaría enormemente a los hogares de menor renta.

Las diferencias en capital humano del padre, tomadas en su conjunto, son el principal determinante del tipo de centro al que envían a sus hijos. Un incremento en el coste de la educación afectaría principalmente a los hogares que tiene un mayor déficit educativo.

Por último, cabe destacar que al aplicar el análisis de componentes principales se observa que, las variaciones en el gasto en educación de los hogares gallegos está influenciado por el precio de estos bienes o servicios, y esto conllevaría a una variación en el comportamiento del gasto de los hogares.

En consecuencia, los criterios de equidad son los más poderosos para justificar la financiación pública en los niveles de educación analizados.

Notas

1. EPF, Encuesta de Presupuestos Familiares, Instituto Nacional de Estadística. Este trabajo ha sido realizado utilizando esta encuesta en su formato magnético o cinta. Los datos expuestos en este trabajo y que no son referenciados por ninguna fuente provienen de esta encuesta. La EPF utilizada en este trabajo fue realizada en el periodo Marzo 1990- Abril 1991.
2. La estructura educacional considerada en este trabajo es la fijada por la Ley Española General de Educación (LGE) 14/1970. 4 de agosto.
3. El conflicto entre la eficiencia y la equidad aparece cuando se alcanza la eficiencia paretiana; para más información ver Blaug M y Moreno, J.L. (1984).
4. Nosotros consideramos el valor dollar/Pts usando el PPPs de 1996.
5. La renta se define como el gasto monetario total de la familia.
6. Al igual que otros trabajos españoles que han utilizado la EPF, ver por ejemplo: Medel et al. (1990), hemos considerado que lo más adecuado es utilizar el gasto medio anual de los hogares como

medida de la renta disponible.

7. Esta tabla, al igual que las siguientes son de elaboración propia a partir de la EPF90/91 en formato cinta magnética.
8. Se ha supuesto que los hogares tienen tan sólo un hijo estudiando en cada uno de los niveles.
9. Dentro de la estructura educativa considerada en este trabajo es necesario establecer una distinción entre la educación infantil y el resto de los niveles educativos analizados. El gasto de la educación infantil no cabe la posibilidad de ser financiado en su totalidad por el sector público por no estar acogido a la enseñanza obligatoria cuando el alumno tiene menos de cuatro años.
10. Por sustentador principal denotamos el cabeza de familia que aporta la mayor contribución económica. Desde ahora nosotros consideraremos como padre al sustentador principal.
11. Se denota la condición socioprofesional del padre por SC.
12. La categoría designada como 1 corresponde con los trabajadores manuales de la industria y los servicios; la categoría 2 corresponde con los trabajadores no manuales en la industria y los servicios; el 3 con los trabajadores autónomos en la industria y los servicios; el 4 con los trabajadores de la agricultura; el 5 con los retirados; y la categoría 6 con otras situaciones profesionales no contempladas en las categorías anteriores.
13. EPA, Encuesta de Población Activa, Instituto Nacional de Estadística de España. Esta encuesta esta considerada como el principal punto de referencia dentro del estado español a la hora de analizar la población activa y sus características.
14. La condición socioeconómica del padre se denota por SEC.
15. La codificación de la EPF ha sido redefinida del siguiente modo: código 1, empresarios agrícolas y directores; código 2, resto de trabajadores agrarios y miembros de cooperativas agrarias; código 3, empresarios y profesionales no agrícolas y trabajadores por cuenta propia no agrarios; código 4, directores, profesionales y jefes administrativos: código 5, resto de personal de servicios y de las fuerzas armadas; código 6, artesanos, trabajadores cualificados y miembros de cooperativas no agrarias; código 7, trabajadores no cualificados no agrarios; código 8, trabajadores no clasificables en los códigos anteriores.
16. Corresponde el código 0: analfabetos, código 1: sin estudios, código 2: educación primaria, código 3: educación secundaria o equivalente, código 4: educación secundaria superior o equivalente, código 5: curso de curso de acceso a la universidad, código 6: formación profesional 1° grado, código 7: formación profesional 2° grado, código 8: educación universitaria de ciclo corto o equivalente, código 9: educación universitaria superior o equivalente.
17. Al referirnos al nivel de estudios del padre estamos considerando el nivel de estudios conseguido o finalizado.
18. Para más información sobre el nivel de estudios conseguido o finalizado se puede consultar: Palafox, J. Default, J. Y Pérez, F. (1995).
19. Existe una gran cantidad de literatura económica en la que se

intenta caracterizar la senda vital de ingresos en función del nivel educativo y de los años de experiencia en el mercado laboral. Para más información ver: Gary Becker (1962): "Investment in Capital Human: A Theorical Analysis".- Journal of Political Economy 70, pp 9-49, Gary Becker (1975): "Human Capital".-University Press, New York. Ben Porath (1967): "The Production of Human Capital and the life Cycle of Earnings".- Journal of Political Economy 75, pp 352-365. and Jacob Mincer (1974) : "Schooling, Experience and Earnings" .- New York: National Bureau of Economic Research. Estos autores han demostrado, dentro de las trayectorias de ingresos de ciclo vital, entre otras conclusiones, que las ganancias de los individuos con mayor formación académica son superiores a los que poseen una formación menor.

REFERENCIAS

Ambler, J (1994) "Who benefits from educational choice? Some evidence from Europe". *Journal of Policy Analysis and Management*, 13, 454-476.

Barnett, W. (1994) "Obstacles and opportunities: same simple economics of school finance reform". *Educational Policy*, 8, 436-452.

Becker, G. (1960) "Underinvestment in college education ?". *American Economic Review*, 50, 346-354.

Becker, G. (1962) "Investment in Capital Human: A Theorical Analysis". *Journal of Political Economy*, 70, 9-49,

Becker, G. (1975) *"Human Capital"*.University Press, New York.

Becker, G. (1985) "Human Capital, Effort, and the Sexual Division of Labor". *Journal of Labor Economics*, 3, S33- S58.

Blair, J. & Stanley, S. (1995) "Quality competition and public shools: further evidence". *Economics of Education Review*, 14, 193-198.

Blaug, M. & Moreno, J.L.(1984*) "Financiación de la Educación Superior en Europa y España"* Siglo XXI. Madrid.

Blaug, M. (1974) "An economic analysis of personal earnings in Thailand". *Economic Development and Cultural Change*, 23, 1-31.

Blaug, M. (1985) "Where are we now in the Economics Education ?" *Economics of Education Review*, 4, 17-28 .

Borland, M. & Howsen, R. (1992) "Studente academic achievement and the degree of market concentration in education". *Economics of Education Review*, 11, 31-39.

Chubb, J. & Moe, T. (1990a) *"Politics, markets and American's*

schools" Washington. The brookings institution.

Chubb, J. & Moe, T. (1990b) "America´s public schools: choice is a panacea, *The Brookings Review*, 8, 4-14.

Coleman, J. & Hoffer, T. (1987) *"Public and private schools: The impact of communities"*. New York. Basic Books.

Conlisk, J. (1971) "A bit of evidence on the income-education-ability interrelation" *Journal of Human Resources*, 6, 358-362

Couch, J. Shugart, W. and Williams, A. (1993) "Private schools enrollment and public school performance". *Public Choice*, 76, 301-312

Cummings, W. & Riddell, A. (Eds.) (1994) "Alternative policies for the finance, control, and delivery of basic education". *International Journal of Educational Research*, 21, 749-828.

Freire, Mª. J. & Salcines, J.V. (1995) "La equidad en la educación: un análisis de la educación post-obligatoria en la Comunidad Autónoma Gallega" *Cuadernos de Economía*, 67, 463-491

Freire, Mª. J. & Salcines, J.V. (1998) "Empleo y Ocupación: Un análisis de las estructuras ocupacionales de Galicia y España" *Documento de Trabajo 1998/7.*. Facultad de CC. Económicas. Universidad de La Coruña.

Friedman, M. (1962) "Capitalism and freedom" Chicago, IL, University of Chicago Press.

Fuchs, V. (1988) *"Women's Quest for Equality"*. Harvard University Press. Cambridge

Glomm, G & Ravikumar, B (1992) "Public versus private investment in human capital: endogenous growth and income inequality". *Journal of Political Economy*, 100, 818-834

Goldhaber, D. (1996) "Public and private high schools: is school choice and answer to the productivity problem?". *Economics of Education Review*, 15, 93-109.

Hanushek,E.(1996) "A more complete picture of school resource policies". *Review of Educational Research*, 66, 397-409.

Henig, J (1994) *"Rethinking school choice: limits of the market metaphor"*. Pricenton University Press.

Hoxby, C. (1996) "Are efficiency and equity in school finance substitutes or complement?". *Journal of Economic Perpectives*, 10, 55-72.

James, E. (1993) "Why do different countries choose a different public-private mix of educational services?" *Journal of Human Resources*, 28, 571-592

Jencks, C. (1970) *"Educational vouchers"* Center for the study of public policy. Cambridge.

Kingdon, G. (1996) "The quality and efficiency of private and public education: a case-study of urban India". *Oxford Bulletin of Economics and Statistics*, 58, 57-82.

Lamdin, D. & Mintrom, M. (1977) "School choice in theory and practice: Taking stock and looking ahead". *Education Economics*, 5, 211-244

Lankford, R. & Wyckoff, J. (1992) "Primary and secondary school choice among public and religious alternatives". *Economics of Education Review*, 11, 317-339.

Lankford, R., Lee, E. and Wyckoff, J. (1995) "An analysis of elementary and secondary school choice". *Journal of Urban Economics*, 38, 236-251.

Levin, H. (1992) "Market approaches to education: vouchers and social choices", *Economics of Education Review*, 11, 279- 285.

Long,J.& Toma, E.(1988) "The determinants of private school attendance". *The Review of Economics and Statistics*, 70, 351-357.

Medel et al. (1990) "La distribución del gasto del estado en educación en España 1981" *Investigaciones Económicas*, XIV, 127-148.

Mincer, J. (1974) *"Schooling, Experience and Earnings"* .New York: National Bureau of Economic Research.

Mincer, J. and Polachek, S. (1974) : "Family investments in Human Capital; Earnings of Women". *Journal of Political Economy*, 82, S76-S108.

Palafox J, Default J, and Pérez, F.(1995) *"Human Capital, Education and Employment"*. Fundación Bancaixa. Valencia. Spain.

Peacock, A. (1983) "Education voucher schemes". Economic Affairs, 3

Picus,L.(1994) "Achieving program equity: are markets the answer? *Educational Policy*, 8, 568-581.

Porath, B. (1967): "The Production of Human Capital and the life Cycle of Earnings".- *Journal of Political Economy*, 75, 352-365.

Psacharopoulos, G. (1980) *"Higher Education in Developing Countries*

: *A Cost Benefit Analysis"*. World Bank, Washington D.C.

Smith, K. & Meier, K. (1995) "Public choice in education: Markets and demand for quality education". *Political Research Quaterly*, 48, 461-478.

Stiglitz, J. (1974) "The demand for education in public and private school systems" *Journal of Public Economics*, 3, 349 -385.

Stiglitz, J. (1988) "La economía del sector público" Ed. Bosch. Barcelona

West, E. (1991) "Public schools and excess burdens". *Economics of Education Review*, 10, 159-169.

Witte, J. (1992) "Private school versus public school achievement: are there findings that should affects the educational choice debate?". *Economics of Education Review*, 11, 371-394.

Autores

Los autores son miembros del Departamento de Análisis Económico de la Universidad de La Coruña (España). Dirección: Facultad de Económicas. Campus da Zapateira, s/n. 15071 La Coruña. España. e-mail: maje@udc.es (Dra. Freire) y jvsc@udc.es (Dr. Salcines)

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
Ann Leavenworth Center
for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North
Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

This article has been retrieved 743 times since April 26, 1999

Education Policy Analysis Archives

Volume 7 Number 16 April 26, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Race and Policy

Ernest R. House
University of Colorado

Abstract

Beliefs about race have played a central role in American history, literature, and education. Racial beliefs are embedded in the national identity in complex and disguised ways. These beliefs attribute presumed character traits to African Americans and other minorities, who are thought of as different in character and ability, especially the ability to govern themselves. These beliefs lead to education policies which separate, differentiate, and mandate different curricula and treatment for minorities, policies justified as being fair and democratic. These beliefs influence not only curriculum content, but how the schools are organized, financed, and administered at a deeper level than is commonly understood.

In his portrayal of the stark inequalities among American schools, Kozol (1991) reported a discussion of race and inequality with white teenagers in a wealthy New York suburb. These teenagers expressed three main beliefs. First, fiscal inequalities among schools don't really matter that much, although loss of funds from their own school to equalize resources with poor schools would be damaging. Second, any form of racial integration would be met by strong resistance in their community, especially by their parents, partly for fear education standards might decline. Third, achieving equity in funding would not make much difference since poor, minority children would fail anyhow. They lack motivation, parental guidance, and a good environment. At this point in the discussion a frustrated boy exclaimed, "When people talk this way ... they're saying that black kids will never learn.... So what it means is—you are writing people off" (Kozol, 1991, p. 130). These beliefs are a product of racism in American society and a major reason it continues. How is it that these students come to hold such beliefs and what difference does it make for education?

For a long time, I have been puzzled by the connection between race and education policy. Despite attempts to ameliorate racism and raise the school performance of minorities, improvements have been modest. Why does racism persist? How do race and policy affect each other? I want to offer a tentative explanation of this connection. First, Americans hold deep-seated beliefs about democracy, equality, and fairness. These beliefs are sincere, I believe. Second, America is a deeply racist country in a particular way. Although many countries harbor racist beliefs, those in America are peculiar in some respects. Third, most "white" Americans don't fully comprehend that their country is racist, nor the extent of that racism, nor how that racism is embedded.

These beliefs result in seemingly contradictory policies. Given these beliefs, I would expect that policies with racial import will often be invisible or disguised as nonracial. That is, the intents and consequences of some policies will be taken to be something other than what they are, even to those who espouse them. This disguise enables the policies to seem fair and democratic, even when the policies have racial overtones. For example, the sociologist William Julius Wilson (1987) contends that Americans will not support policies that are believed to benefit minorities primarily. They will support social security since that is seen as benefiting all Americans, including minorities. But they will not support welfare programs, which are perceived as helping mainly minorities.

I have a corollary to Wilson's thesis. Americans will support policies that are harmful to minorities that they would not tolerate if those same policies were applied to majority populations. In education, for example, Americans are strongly in favor of retention—retaining students at the same grade level for another year—even though the research evidence overwhelmingly shows strong negative effects on the students retained. Retention programs are applied massively to minorities in large cities, but not to majority populations. Yet retention does not appear on the surface to have racial implications. Other education policies that appear to have little to do with race also severely disadvantage minorities, including how schools are financed, how schools are organized, how standardized tests are used, and how students are grouped. In other words, we have organized ourselves educationally in ways to disadvantage minorities, even while maintaining appearances of equality in such matters. No other developed countries have organized their educational systems in this fashion.

I would call our system one of institutional racism, racism not recognized by those participating in it because the way the institutions function seems normal to those growing up in them. Individuals need have no hostile racist thoughts, only adherence to the system as it exists. I would not say that the way we have organized ourselves educationally is determined solely by racial concerns, only that such concerns are one powerful causal factor. Other economic and social factors also have had strong effects. Now these are strong claims, outrageous claims, some will say. How can they be substantiated? No single attempt of this length could conceivably deal definitively with the problem of American racism. In what remains of this attempt, I would like to explore explanations as to how racism affects education policy: first, how race has been constructed and used in America historically; second, how race is embedded in American identity; and, third, how these beliefs play out in education policies.

Our National Identity

Reluctantly, I have come to believe racism is deeply embedded within the national identity itself, built into the American character by history and experience. Jacobson (1998) traced in detail how the concepts of "race" and "whiteness" have been socially constructed and employed in the US. In his history of the concept of "whiteness" he showed how immigrants were assimilated through redefinitions of what "whiteness" meant. What were seen as separate "races" of people gradually came to be socially constructed as "white ethnics," but not without struggle.

Early colonial Americans defined themselves as "white" and "free" in contrast to those who were not, especially slaves and Native Americans. As early settlers escaped the class systems of Europe, they redefined themselves along racial lines, yet another hierarchy of human worth. The first colonists were mostly English, and they held strong beliefs of Anglo-Saxon racial superiority, beliefs that the English displayed throughout their world colonies.

The notion of different "races" with different characteristics embedded in nature presumed many such "races"—Celts, Teutons, Poles, Swedes, Hebrews, Turks, and so on. Race was assumed to be hereditary, fixed, and closely associated with character traits. Knowing people's race was a shorthand way of knowing about their character.

Race was also connected to one particularly important attribute: The Anglo-Saxons saw themselves as uniquely equipped to govern by reason of their natural superiority, whereas other races were less well endowed. The idea that the ability to govern endowed its possessors with rights over others had a home in the British upper classes, and this idea was generalized to foreign populations that the English contrasted with themselves, especially those colonized. Racism so conceived was "a theory of who is who, of who belongs and who does not, of who deserves what and who is capable of what" (Jacobson, 1998, p. 6).

Belief in their natural, God-given ability to govern became a primary justification for appropriating land and resources that others could not manage properly, i.e., to take land and resources away from others so the resources could be used productively. For the British these others included colonial peoples all over the world. For Americans these others included African Americans, Native Americans, Mexicans, and Filipinos, all of whom possessed contested property at one time or another.

In the beginning, being white meant being Anglo-Saxon. All others were of different races. Founders of the nation (southern gentlemen planters) wrote restrictions on these others into law, including their rights to hold property, become citizens, enter the country, and even be owned by whites. Race was closely associated in the courts with property rights. The very idea of "providing for the common defense" in a country of slaves and frontier settlements was racially imbued.

The 1790 naturalization law was fiercely exclusionary, and the early statutes suggested that "whiteness" was the criterion for full citizenship. Soon, this belief was challenged by the arrival of assorted immigrants. The need for cheap labor for settlement and economic development opened up immigration to all kinds of "races." Large numbers of immigrants arrived, such as the Germans and Irish. These people were seen as different races ("Teutons" and "Celts") with their distinct physical features signifying different character traits. The Irish were seen as a physically dark race which was certainly not up to self-governance (as in colonized Ireland). Their assimilation was problematic. Over decades each new immigrant group struggled to be defined in the "white" mainstream, and for what that portended for social and economic advantage. Eventually, most groups were redefined as "white ethnics" who were different in "culture" but not necessarily race. In other words, their ethnicity was derived from their culture and might be remedial, as opposed to their "race," which was not. Just as the research of earlier times buttressed theories of "race" so too did later research support theories of "ethnicity" when their time came. However, the assimilation of the "ethnics" (episodes portrayed as American success stories) deepened the chasm between those "white" and those not. If Jews, Syrians, and Turks were "white," where did this leave African Americans?

In fact, the reason all these immigrants were considered white ultimately was that they were contrasted to African Americans, who were never considered white. The immigrants gained whiteness by African Americans being excluded. And it was to the advantage of immigrants to adopt beliefs and positions which were overtly racist in order to contrast themselves to African Americans. Hence, the racist dichotomy enabled the ethnics to become "Americanized" at the cost and through the exclusion of others. What was conceived literally as many separate races early in American history was reduced to a white-black dichotomy in the 20th century, complicated by a large influx of Latinos.

Hence, in a sense the "white"—African American racial relationship is at the center of American national identity, though it is not recognized as such. Public policy swirls around it much of the time, even when its influence is not recognized explicitly. Typically, African Americans are portrayed as an unfortunate group which just happens to be marginal, a minority which has never quite made it for a number of reasons, (again) having to do with their character traits, e.g., lack of skills, lack of self-discipline, impoverished culture, low natural intelligence—and most certainly lack of ability to govern themselves.

In a sense African Americans were never marginal. They have been central to the formation of the American national identity in that they provided a primary touchstone against which "whites" defined themselves, though other "uncivilized" peoples such as Native Americans also played such a role. "Racism now appears not anomalous to the working of American democracy but fundamental to it" (Jacobson, 1998, p. 12).

The Formative Process

The nature of this complex relationship is difficult to untangle by those shaped by it, just as it is difficult for males to understand how they construct their own identities by defining females in certain ways. A glimpse into the complexities of national identity is provided by Nobel Laureate Toni Morrison's examination of the national literature.

In *Playing in the Dark: Whiteness and the Literary Imagination* (1992), Morrison noted that it is often the literary image that is not presented or which is presented obliquely that reveals distorted and repressed thoughts and feelings about race.

> Race has become metaphorical—a way of referring to and disguising forces, events, classes, and expressions of social decay and economic division far more threatening to the body politic than biological race ever was .... It seems that it has a utility far beyond economy, beyond the sequestering of classes from one another, and has assumed a metaphorical life so completely embedded in daily discourse that it is perhaps more necessary and more on display than ever before.... I remain convinced that the metaphorical and metaphysical uses of race occupy definitive places in American literature, in the "national character"... (Morrison, 1992, p. 63-64).

How is it possible that Americans could harbor and repress racial images, disguising them even to themselves? The early colonists located themselves in a vast continent far from European civilization, besieged by what they saw as threatening Nature and savage forces. Escape from the past was one goal, freedom another.

In this environment unfree slaves served as surrogate selves for the colonists' literary meditations on human freedom, their terror of being European outcasts, their dread of failure, their fears of powerlessness, of nature without limits, of loneliness, and of internal aggression. "What arose out of collective needs to allay fears and to rationalize external exploitation was an American Africanism—a fabricated brew of darkness, otherness, alarm, and desire that is uniquely American" (Morrison, 1992, p. 38). Nothing highlighted freedom like slavery.

The explicit focus of much early American literature was the "architecture of the new white man" (Morrison, 1992, p. 15), the deliberate construction of the new American. This new image was created by finding differences with Europeans, as Emerson did, but also in contradistinction to those closer—slaves, Native Americans, and Latinos. The flight from the Old World to the New World was a flight from oppression and limitation to freedom and responsibility, but in a peculiar way.

Morrison illustrates this developing national identity in the character of an early planter slave-owner (taken from Bernard Bailyn's (1986) *Voyagers to the West*, an investigation of settlers becoming Americans). William Dunbar was a young Scottish aristocrat educated by tutors and in math, astronomy, and belle-lettres at the University of Aberdeen. He became a London intellectual, an exemplar of Enlightenment thinking. But after moving to the New World he suppressed a slave conspiracy on his Mississippi plantation in 1776. He was astounded at his slaves' ingratitude after he had treated them so well, and he administered two runaways 500 lashes on five separate occasions. Morrison cites Bailyn's assessment of Dunbar's character transformation: "...feeling a sense of authority and autonomy he had not known before, a force that flowed from his absolute control over the lives of others, he emerged a distinctive new man, a borderland gentleman, a man of property in a raw, half- savage world" (quoted in Morrison, p. 42). Morrison takes these traits as prototypical of the new white American male, traits valorized and inscribed as themes in our national literature.

The sense of autonomy was transformed into themes of American "individualism." Newness was transformed into American "innocence." Distinctiveness was transformed into "difference" and strategies for maintaining difference. Authority and absolute power over others were transformed into a conquering romantic heroism and masculine virility—and also into the moral problematics of exercising such power over others. According to Morrison, the major characteristics of our national literature—individualism, masculinity, social engagement versus historical isolation, acute moral problematics, and themes of innocence coupled with figurations of death and hell are responses to an Africanist presence—guilt, violence, alienation, power, freedom. American coherence of identity was organized in part through a distancing Africanism.

Hence, race functioned as a metaphor for the construction of American identity. "Deep within the word 'American' is its association with race." (Morrison, 1992, p.47). "What was distinctive in the New was, first of all, its claim to freedom, and, second, the presence of the unfree in the heart of the democratic experiment...." (p. 48). The image of bound, suppressed, and repressed darkness became objectified as an Africanist persona. "Africanism is the vehicle by which the American self knows itself as not enslaved, but free; not repulsive but desirable; not helpless, but licensed and powerful; not history- less, but historical; not damned, but innocent; not a blind accident of evolution, but a progressive fulfillment of destiny" (Morrison, 1992, p. 52). According to Morrison, racial themes are embodied throughout American literature, in spite of claims that American literature is raceless. In denying racism these matters were discussed in disguised ways. The consequence was a master narrative that spoke for Africans but not of them. One theme is the view of the Africanist character as enabler, somehow pulling free of claims of retribution, as in *Huckleberry Finn*. Another is how African American characters are used to limn out and enforce whiteness and how they are used strategically to define the goals and qualities of whites, e.g., "to control projections of anarchy with the disciplinary apparatus of punishment and largess." (p. 53).

> My suggestion that Africanism has come to have a metaphysical necessity should in no way be understood to imply that it has lost its ideological utility. There is still much ill-gotten gain to reap from rationalizing power grabs and clutches with inferences of inferiority and the ranking of differences. There is still much national solace in continuing dreams of democratic egalitarianism available by hiding class conflict, rage, and impotence in figurations of race. Freedom (to move, to earn, to learn, to be allied with a powerful center, to narrate the world) can be relished more deeply in a cheek-by-jowl existence with the bound and unfree, the economically oppressed, the marginalized, the silenced. (Morrison, 1992, p. 63-64)

Morrison notes that the presence of black people, along with gender and family, are inherent in the earliest lessons children are taught regarding their distinctness. And that "It is no accident and no mistake that immigrant populations (and much immigrant literature) understood their "Americaness" as an opposition to the resident black population" (p. 47).

Education Policy

If such beliefs and images are built into the national identity, what ramifications would we expect them to have for educational policy? Presumably policies would be based on maintaining separation, maintaining differences based on racist images of the Africanist persona's presumed character traits. Such policies would be consistent with race themes that reoccur in American history and literature.

The teenagers in the New York suburb believed that fiscal inequalities don't matter, that racial integration was hopeless since it would be resisted, and that equity would not help minority students since they are beyond help. *They* lack motivation, parental care, and a healthy environment. *They* are throwaways, as the boy declared. The historical image of African Americans is recognizable in the beliefs of these white teenagers.

If such majority beliefs guide policies, what would the policies be like? First, the policies would keep African Americans separate and distinct from "white" Americans. Second, they would be based on the presumed character traits of African Americans, traits that differentiate them from whites; presumably, traits which mark them as inferior. Their education would be different from whites in important ways because of these traits. Third, since they are not capable of self-governance they cannot be in control of their own education. Fourth, since African Americans have limited potential, investment in their education should be modest: no need to over invest since what can emerge is limited. Fifth, such policies would attempt to appear fair, to conform to the American Creed (e.g., Mrydal, 1964).

In fact, these are the types of policies that have characterized the treatment of African Americans: exclusion, differentiated curricula and treatment based on presumed inferior character traits, control by others, and modest resource commitment, all justified as being fair. Similar policies have applied to other minorities as well—to Native Americans, Latinos, and others, though each group has its own unique history and experience. I will illustrate these themes with three examples (though the available examples are beyond cataloguing): the education of African Americans in the South after the Civl War, desegregation in Chicago in 1967, and retention ("failing" grades) in Chicago in 1998.

The Education of Blacks in the South, 1860-1935

Anderson (1988) documented the history of black education in the South between 1860 and 1935. Contrary to popular belief, the ex-slaves were very keen on establishing an education for themselves and their children since they saw education as the way to attain freedom and escape poverty. Once freed, they immediately set up schools for themselves. However, the southern planters, still in control, resisted public education since it was a threat to their forced agricultural labor supply, which depended on child labor. Furthermore, educated African Americans were a threat to the myth of black inferiority since poor whites in the South had no public education themselves. By contrast, northern industrialists saw basic education as aiding worker efficiency—provided it was focused on vocational training, the only kind of work African Americans were capable of. While the ex-slaves wanted teachers, managers, and businessmen trained academically to be leaders, the whites wanted vocational training for them. Hence, southern and northern whites supported the Hampton Institute and Tuskegee model of training. Both the control and goals of African American education were imposed by whites. Hampton Institute trained teachers, but the curriculum was focused on hard manual labor and on instilling the proper political and moral attitudes so that these teachers could train students for their proper role in society—disciplined, low-skill manual laborers. The founder and principal, a northerner, saw African American voters as immoral and irresponsible; they should not participate in public life. Black masses were weak, and black leaders were "ignorant, immoral preachers or selfish politicians" (Anderson, 1988, p. 38). These ideas were embodied in teacher training throughout the South by philanthropist money and influence.

By the early 20th century working class whites finally had public elementary schools in the South—from which African Americans were excluded. During this period, African Americans built their own common schools with their own money and with contributions from local whites and philanthropists. Again, whites insisted that African American high schools be vocational schools to train workers for jobs that African Americans held, not for jobs held by whites. However, the Great Depression threw white workers into competition over previously black jobs, and support for even vocational high schools ceased. Anderson ends his book by noting the irony of scholarship which attributes low African American educational attainment to initial differences in cultural orientation towards education, and studies which blame black dialect, oral traditions, and cultural separatism for preventing school success. Since Emancipation the ex-slaves had done everything they could to educate their children, and they were impeded vigorously at every turn by whites determined to keep them in their place.

This history embodies key policy themes of African American education over decades: exclusion and segregation, based on images of African Americans as not very capable intellectually, as being immoral, and as being not qualified for political governance. Control over education was exercised by whites and resulted in differentiated curricula which trained African Americans for lower positions in society.

Chicago Desegregation, 1967

Peterson's (1976) analysis of desegregation of the Chicago schools examines policies in a different time and place. In 1967, Chicago's schools were highly segregated. African American students were 52.3 percent of the public school population, but only 4.7 percent of African American students were in schools more than 50 percent white. Only 28 percent of white students were in schools more than 5 percent African American.

The busing plan developed by the Chicago superintendent of schools was modest: to bus 1,000 African American students to schools in two neighborhoods with white student populations of 10,000. This was part of a larger plan of magnet schools, educational parks, and financial incentives to promote racial balance. Initially, the busing plan was approved by the school board "in principle." But it was never implemented. African American leaders saw the plan as too modest. Whites in the affected neighborhoods were outraged by even this much integration, and they mobilized into an uncompromising resistance.

Although those supporting the plan were able to gain the endorsements of civil rights groups, newspapers, the teachers union, and the Catholic church, opposing whites were able to enlist the support of the powerful political machine itself, especially US Representative Pucinski, who represented ethnic neighborhoods. Demonstrations, picketing, and intense lobbying were part of the campaign against the busing of African American students. Peterson (1976) labels the decision process of the school board "ideological bargaining." Opponents took "principled" stands on issues which they were unwilling to compromise. They used public channels of communication to politicize the issues before the general public, so that outside parties became involved. The school board members themselves voted the beliefs with which they had started. Only one board member was moved to change sides, though his vote was critical in neutering the busing policy.

The principle on which the opposition to busing rested was that of "neighborhood schools." Those in the neighborhood should be able to send their children to nearby schools and control those schools. Others should stay out. Mayor Richard Daley left it to the school board to settle, but he indicated he favored neighborhood schools so that parents could "...see their children home for lunch, and discuss what happened in school with them" (quoted in Peterson, 1976, p. 160).

The overall guiding policy for the city was "racial balance," which meant that policies should keep the racial distribution in the city as it was. Otherwise, it was feared, whites would move to the suburbs, thus making the city increasingly minority. Neighborhood schools were seen as facilitating racial balance. But when plans were made to maintain this racial balance by introducing African American students into a few schools in modest numbers, the white population wanted none of it. Something similar happened in other cities. Again, the same features apply as for policies in the South: exclusion from the mainstream, vigorous opposition by whites, and control of African American education by whites. Whether students were subjected to a differentiated curriculum is not discernible from Peterson's analysis, nor are we told the presumed traits of African Americans. Finally, this neighborhood policy was seen as fair since any people could have their own neighborhood schools in principle.

Chicago Retention, 1998

Thirty years later, Chicago schools were more than 80 percent minority. The racial balance policies had not kept whites in the city, nor white children in the public schools. Whites had fled the city or put their children in private schools. What also had occurred was a radical restructuring of the schools, a decentralization to the local school forced by the Illinois legislature, which had become frustrated by the inability of the Chicago schools to stay within their budget. Acting on the ideas of Designs for Change, an educational activist group, each school became governed by a local school council. This decentralization was an attempt to wrest control of the local schools away from the central bureaucracy, an idea that would appear to conform closely to the "neighborhood school" principle, the presumed basis for blunting desegregation years before.

However, what happened in the late 1990s was that Mayor Richard M. Daley, son of the former mayor (Richard J. Daley), took control back from the local school councils. New legislation gave the Mayor control over the schools. Daley recentralized the schools and appointed his aide as head. The major educational reform policy became "retention," making students repeat a grade if they did not attain a specific cut-off score on a standardized test. More than 12,000 elementary students were retained in grades 3, 6, and 8 in 1996 and 1997. Schools with the highest rates of failure were 69% African American, 27% Latino, 3% white and 94% low income. In the 1997-1998 school year, Chicago's Transition Centers (special high schools for the retained) enrolled 929 African Americans, 330 Latinos, and 34 whites.

Furthermore, a new curriculum consisting of more than 4,000 lesson plans based on the standardized achievement tests was implemented, and teachers were monitored to ensure that they taught these lessons. Such practices are far removed from the "neighborhood schools" principle. Again, the idea of minorities running their own schools did not seem viable to white authorities.

Retaining students has proved to be one of the worst educational practices. The research evidence against it is overwhelming. Retained students do not achieve more academically and are much more likely to drop out of school later (Shepard and Smith, 1989). Often they become stigmatized by their failure. In one study, flunking a grade increased the chances of an African American male dropping out of school by 38 percent and a white female by 17 percent, all other factors being equal (Grissom and Shepard, 1989). Those being retained are mostly minority males.

New York City had a similar retention program in the 1980s in which the school district held back twenty-five percent of students in fourth and seventh grades. The program resulted in no greater academic gain and in higher drop- out rates for those students retained, who were more than 80 percent minority. Nonetheless, in the late 1990s retention was the popular program of choice for minority students in large cities across the country. The Chicago program was praised as exemplary by President Clinton in his 1999 State of the Union address, and Clinton proposed similar retention policies all across the country enforced by federal sanctions.

Massive retention policies display features similar to other educational policies for minorities. First, students are excluded from the mainstream. Second, they are taught a differentiated curriculum. In the case of Chicago, the curriculum is based on the standardized test being administered to retain them. Unfortunately, teaching a particular test often does not generalize to other forms of academic achievement (Linn, 1998). The differentiated curriculum is based on the presumed characteristics of those retained, i.e., their low achievement levels and it is justified by objective test scores. Third, the process is controlled by whites running the system.

While William Julius Wilson suggests that Americans will not support policies seen to benefit minorities primarily, Americans do seem to support policies which are detrimental to minorities. They would not support massive retention if it were applied to majority populations. The Chicago suburbs fail less than one percent of their students. If one looked at the explicit rationales for Chicago's neighborhood schools and Chicago's later retention program, one would be puzzled by apparent contradictions. Yet both sets of policies are fully consistent with the underlying beliefs about minorities and with the education policies applied to minorities.

The Current Educational System

Although American racism has improved discernibly over the past several decades and minorities have more rights than they used to, basic beliefs remain in place. In a national survey of racial attitudes, Schuman, Steeh, and Bobo (1985) concluded:

> Relations between blacks and whites in this country have been based since the beginning on a conception of two socially distinct groups defined largely in terms of physical characteristics....There is no evidence that this virtually absolute differentiation of the American population has been reduced by any of the changes of the past four decades. Indeed, it may even have increased.... America is not much more color-blind today than it ever was.... (Schuman, Steeh, and Bobo, 1985, p. 201).

Carmines and Stimson (1989) demonstrated how race became the major domestic partisan issue in American politics after 1964. It was not a partisan issue before because whites were agreed about the place of African Americans: African Americans were disenfranchised. However, economic change led to the development of a new African American leadership class and the Civil Rights movement. Beginning with the 1964 Johnson and Goldwater Presidential contest, race became a major issue separating Democrats and Republicans, albeit in disguised form. The issue was expressed in the form of whether the federal government should take steps to alleviate racial problems or whether this task was beyond its proper scope ("The government cannot solve all our social problems".)

It would be amazing if such deep-seated beliefs did not affect education policy. In fact, I believe that education policy has been profoundly shaped by these beliefs. Americans have defined their educational system in such a way as to ensure that African Americans (and often other minorities) are treated in an exclusionary way—which is to say that they are saddled with an education which is inferior, and this inferior education contributes to whites seeing them as having undesirable attributes and as being unable to govern themselves.

Here are some education policies that produce these outcomes, even though the policies may have been implemented from motives other than racial ones:

- School organization—Schools are organized into separate governing entities which effectively encapsulate and isolate

African Americans and other minorities in large numbers. Large school districts are overwhelmingly minority, and these districts are rarely controlled by the minorities. White politicians and policymakers effectively control the fate of minority children.

- School finance—Schools are financed through local sources in large part, which ensures resources are hugely unequally distributed among school districts, with minority schools faring particularly poorly financially. Kozol (1991) has illustrated these inequalities dramatically.
- Curricula—Schools with large numbers of minorities receive educational curricula unlike those in majority schools. These curricula are geared to "slow" students and are repetitious and boring, based on presumed lesser abilities of minority students (Levin, 1993).
- Ability grouping—Within local schools students are separated into differentied groups, and the low-ability groups contain disproportionate numbers of minorities (if there are any) who receive differentiated curricula (Oakes, 1985, Wheelock, 1992).
- Retention—Flunking students on a massive scale is done in districts with large numbers of minorities. Being retained does not increase learning but rather significantly increases the student's chances of dropping out of school eventually (Shepard and Smith, 1989).
- Testing—Achievement tests legitimate these activities to make differentiation appear fair. Students are selected for ability groups and retained in grade by test scores. If students are selected by objective assessments, how can the practices be unfair? The way tests are used contributes much towards legitimating racial policies, even as test publishers publicly deplore such uses (Haney, 1993).

Each of these educational practices can be explained and justified without reference to race. Local schools, local financing, standardized testing, ability grouping, retention, and differentiated curricula can appear to have little to do with race. They could stem from American beliefs in autonomy, freedom, and meritocracy. But a prolonged examination of how these policies function together—and who they are applied to—reveals that the policies effectively segregate, differentiate, and provide minorities with an inferior education. The operation of the system as a whole has racial consequences even if those administering it do not have that in mind.

No other developed countries incorporate all these practices in this fashion. Most countries have common curricula, much more equitable funding, and central distribution of resources. Few countries employ ability grouping at the primary levels, and no developed countries that I know employ massive retention based on single test score cutoffs.

The structure of our entire educational system has been strongly influenced by the beliefs that people hold about minorities, particularly about African Americans. No doubt there are other contributing causes for these policies as well. No policy ever results from a single cause.

Nonetheless, American education as a whole functions as a racist system, whatever its virtues might be.

References

Anderson, J. D. (1988). *The education of blacks in the south, 1860-1935.* Chapel Hill, NC: University of North Carolina Press.

Bailyn, B. (1986). *Voyagers to the west: A passage in the peopling of America on the eve of revolution.* New York: Alfred A. Knopf.

Carmines, E. G. and Stimson, J. A. (1989). *Issue evolution: Race the transformation of American politics.* Princeton, NJ: Princeton University Press.

Grissom, J. B. and Shepard, L. A. (1989). Repeating and dropping out of school. In Shepard, L. A. and Smith, M. L. (Eds). *Flunking grades: Research and policies on retention.* London: Falmer. 34-63.

Haney, W. (1993). Testing and minorities. In L. Weiss and M. Fine (Eds.). *Beyond silenced voices.* Albany, NY: State University of New York Press. 45-73.

Kozol, J. (1991). *Savage inequalities.* New York: Crown.

Levin, H. L. (1993). The economics of at-risk students. In E. P. Hoffman (Ed.), *Essays on the economics of education.* Kalamazoo, MI: W. E. Upjohn Institute for Employment Research. 11-33.

Linn, R. L. (1998). Assessments and accountability. Center for Research on Evaluation, Standards, and Student Testing, Boulder: University of Colorado.

Morrison, T. (1992). *Playing in the dark: Whiteness and the literary imagination.* Cambridge, MA: Harvard University Press.

Myrdal, G. (1964). *An American dilemma.* 2 volumes. New York: McGraw-Hill.

Jacobson, M. F. (1998). *Whiteness of a different color: European immigrants and the alchemy of race.* Cambridge, MA: Harvard University Press.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CN: Yale University Press.

Peterson, P. E. (1976). *School politics, Chicago style.* Chicago: University of Chicago Press.

Schuman, H., Steeh, C., Bobo, L. (1985). *Racial attitudes in America.* Cambridge, MA: Harvard University Press.

Shepard, L. A. and Smith, M. L. (Eds.). (1989). *Flunking grades: Research and policies on retention.* London: Falmer.

Wheelock, A. (1992). *Crossing the tracks.* New York: New Press.

Wilson, W. J. (1987). *The truly disadvantaged.* Chicago: University of Chicago Press.

About the Author

Ernest R. House

Email: Ernie.House@Colorado.edu

Ernest R. House is a professor in the School of Education at the University of Colorado at Boulder. Previously, he was at the Center for Instructional Research and Curriculum Evaluation (CIRCE) at the University of Illinois, Urbana-Champaign.He has been a visiting scholar at UCLA, Harvard, and New Mexico, as well as in England, Australia, Spain, Sweden, Austria, and Chile. His primary interests are evaluation and policy analysis.

Books authored include *Evaluating with Validity* (1980), *Jesse Jackson and the Politics of Charisma* (1988), *Professional Evaluation: Social Impact and Political Consequences* (1993). His most recent book, *Schools for Sale* was published in 1998.

He is the 1989 recipient of the Harold E. Lasswell Prize presented by Policy Sciences and the 1990 recipient of the Paul F. Lazarsfeld Award for Evaluation Theory, presented by the American Evaluation Association. He was editor of *New Directions in Program Evaluation* (1982 to 1985) and columnist for *Evaluation Practice* (1984-89). Major studies he has directed or particpated in include evaluation of the Illinois Gifted Program for the Illinois legislature (1968-1972), assessment of the Michigan Accountability Program for the National Education Association (1974), critique of the National Follow Through Evaluation for the Ford Foundation (1977), audit of the Promotional Gates Program evaluation for the Mayor's Office in New York City (1981), assessment of environmental education policies in Europe for OECD (1992), and evaluation of science, engineering, and technology education programs across federal departments for the Federal Coordinating Council for Science, Engineering, and Technology in Washington (1993).

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs- Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

David D. Williams
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa- DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

Este artículo ha sido consultado 168 veces desde el 29 de avril de 1999

Education Policy Analysis Archives

Volume 7 Number 17 **Avril 29, 1999** **ISSN 1068-2341**

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ
85287-0211

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

¿Autonomía por decreto?
Paradojas en la redefinición del trabajo del profesorado

José Contreras Domingo
Universitat de Barcelona

Abstract

En el actual ambiente de reformas educativas que estamos viviendo, la autonomía del profesorado y de los centros educativos está siendo una referencia habitual. Presentada como devolución de competencias al profesorado, en realidad, la autonomía, tal y como viene siendo propiciada y argumentada en las nuevas políticas educativas, no puede ser considerada como una simple ganancia de atribuciones profesionales. Cuando menos supone una paradoja: la que representa que ésta venga obligada. Pero hay otras. El presente texto está organizado en cuatro secciones. En la primera se exponen diversos factores -que van desde las nuevas formas de organización de los servicios sociales y su relación con el estado, hasta los actuales patrones de producción y consumo, pasando por el papel de legitimación que desempeña la clase académica- que están dando lugar a la retórica de la autonomía. En la segunda, se describen algunos de los cambios que se vienen produciendo en la institución escolar y que están proporcionando un sentido muy especial a la autonomía escolar y profesional. En la tercera sección se analizan las paradojas que acompañan a las políticas actuales de autonomía escolar. En la cuarta, por fin, se extraen algunas conclusiones de los análisis precedentes y se sugieren algunos elementos que deberían tenerse en cuenta para recuperar otro significado político y pedagógico para la autonomía.

Abstract

In the present atmosphere of education reform in Spain, autonomy of the faculty and the educational institutions is continually referred to. "Autonomy" as it is currently used in this context poses a paradox: autonomy as obligation. Not just this paradox but others as well are discussed. The present article is organized in four sections. In the first section, the article analyzes a set of factors (from the new organizational arrangments of social services and their relationship with the State to the current patterns of production and consumption) present in the autonomy rethoric. In the second section the author describes the main changes in the school as institution, changes that are affecting academic and professional autonomy. In the third section, the paradoxes surrounding present policies of scholastic autonomy are analyzed. In the fourth, suggestions are made for recovering the political and pedagogical meaning of autonomy.

1. Los nuevos discursos: las paradojas de las retóricas

del cambio... y de los académicos hablando sobre la autonomía del profesorado

Quizás nunca como ahora se había vivido internacionalmente un ambiente de reformas escolares simultáneas. Ello significa no sólo que en la actualidad la mayoría de los países están preocupados por sus sistemas educativos. Significa también, que *la reforma*, o *el cambio* de la educación (con sus diferentes conceptualizaciones y estrategias) se ha convertido en preocupación, en tema, y en justificación de intervenciones continuas sobre el sistema educativo. Ha invadido el ambiente de lo escolar como un rasgo definitorio: lo estable es ya la reforma. En España, por ejemplo, tras varios años de experimentación, se aprobó la actual ley educativa (la LOGSE) en 1990, y hoy, tras ocho años, seguimos hablando de "la Reforma", como si fuera algo que está ocurriendo ahora, como una novedad todavía por culminarse. (Note 1) En este ambiente internacional de reformas, aunque éstas puede que sean muy variadas de un país a otro, sí parecen obedecer, sin embargo, a pretensiones similares. Son políticas de cambio que se basan en supuestos semejantes de reacomodación de la educación a nuevas exigencias internas y externas. Aunque con traducciones variadas, adaptadas a las distintas peculiaridades de cada país, con sus diferentes coyunturas políticas y económicas, con sus tradiciones y trayectorias diferenciadas en cada sistema educativo, social, político y cultural, todas ellas parecen compartir los mismos principios de una filosofía de fondo y semejantes sistemas de legitimación (Ginsburg, 1991).

Normalmente, la mayoría de estas nuevas políticas educativas se han referido al curriculum, a la organización y gestión de los centros educativos, a las relaciones de éstos con su entorno, y también al propio trabajo del profesorado, que se ha visto redefinido, bien como producto de políticas expresas dirigidas a este fin, bien como consecuencia de las anteriores. (Note 2) Todos estos cambios están siendo especialmente profundos, porque no se limitan a ser una cuestión de contenido pedagógico, como tampoco son sólo cambios en el interior del sistema educativo. Por ejemplo, no es sólo que se hayan defendido nuevos contenidos curriculares para la educación escolar, sino que, en cada país de distinta forma y a partir de diferentes posiciones de partida, (Note 3) lo que se ha establecido es una nueva relación entre el estado, los centros educativos y los enseñantes en la forma de decidir y controlar el curriculum, modificándose las competencias respectivas. Y la trascendencia especial de estos cambios se encuentra en que, lo que se está transformando, en cada país con sus traducciones y estrategias particulares, son las relaciones entre el estado, los servicios públicos y sus usuarios, transformaciones que vienen determinadas por un cambio en la forma del estado, y en la forma de la sociedad civil (Ball, 1997; Whitty, et al., 1998).

Básicamente, podemos cifrar el carácter de estos cambios generales en la tendencia creciente a pasar de un sistema en el que el estado era el garante de los servicios públicos, fijando sus propósitos, financiándolos, planificándolos y administrándolos mediante su organización profesional-burocrática y ofreciéndolos en condiciones semejantes a toda la población, a otro sistema en el que el estado tiende a reducir su función a la fijación de objetivos, dejando a los centros y organismos proveedores de los servicios un mayor margen de planificación y regulación interna para alcanzarlos. Esta reducción del estado de sus competencias planificadoras y administrativas para cederlas a los proveedores de los servicios se ha producido utilizando como factor de motivación la capacidad de variación en las formas de provisión de los servicios, en competitividad con otros proveedores.

Estos cambios no aparecen al margen de la introducción de modificaciones sustanciales, como filosofía y como práctica, en muchos ámbitos de la vida productiva y social en general. Ideas y prácticas tomadas de los nuevos modos de producción post-fordista (Robertson, 1996; 1997) hacen su aparición aquí. Es el caso, por ejemplo de muchas empresas en las que se está introduciendo la flexibilidad en la producción, poniendo en relación la iniciativa del trabajador y la adaptación al consumidor. Los llamados "círculos de calidad", en los que se reconoce la capacidad de iniciativa del trabajador en la adaptación de los procesos de producción son un ejemplo de esto. Los sistemas homogéneos, en los que la "cadena de montaje" era el prototipo de la racionalidad y la productividad en otros tiempos, son ahora cuestionados, en beneficio de la capacidad de decisión de los trabajadores en su trabajo (Allen, et al., 1992). Lo que parece estarse imitando en los servicios públicos son formas de gestión y de producción semejantes.

Las formas de producción flexible, y por tanto también, la filosofía de los servicios públicos como unidades no homogéneas, parten precisamente de la idea de la diversidad social, de las diferentes demandas y necesidades a las que sistema productivo y servicios deben atender de forma diferenciada. Este reconocimiento de la diversidad social tiene varios efectos. De una parte, puede reconocerse como la atención al derecho a elegir, el reconocimiento de las diferencias y la adaptación a las exigencias. Pero de otra, el propio derecho de elección se ha convertido en una obligación. Cuando los productos de consumo se diversifican, el cliente se ve obligado a sopesar entre la diversidad de oferta, y a elegir entre ellas. Hasta en los productos de consumo más cotidianos, la compra se está convirtiendo en una tarea cada vez más compleja. Ya no podemos elegir sólo la leche más barata, o la de nuestra marca preferida o de confianza; ahora tenemos que decidir si la queremos más o menos desnatada, o si la preferimos con calcio o con vitaminas añadidas; además de que, no podemos elegir ya la leche al margen de la invasión de noticias y teorías en competición sobre las ventajas e inconvenientes de este producto para un tipo u otro de dieta. Si esto es así en la construcción de nuestra subjetividad como consumidores de productos alimenticios, mucho más puede llegar a serlo en nuestra elección de ofertas variadas de servicios, en los que la diferencia tiene claros componentes respecto al significado o consecuencias en nuestro estatus o en nuestras oportunidades sociales (Bowe et al., 1994). (Note 4)

Una consecuencia importante de estos cambios, que ya comienzan a parecer inevitables es que se empieza a transformar la imagen de la ciudadanía, acentuando cada vez más su función como consumidora y electora racional. Así, de la misma forma que los trabajadores parecen obtener una mayor capacidad en las decisiones sobre su trabajo, igualmente, los ciudadanos parecen también recuperar una mayor capacidad en la decisión sobre lo que quieren. En relación a los servicios públicos, esta filosofía parece referirse a que el estado disminuye su papel como regulador de los servicios en su provisión a los ciudadanos, para que sean las relaciones de oferta y demanda las que encuentren su equilibrio directamente, o con menos intermediarios.

Autonomía (de los servicios públicos y de los ciudadanos), descentralización, desregulación y devolución del poder del estado a la sociedad son, parece, términos que designan y resumen en sus máximos valores y aspiraciones todas estas transformaciones. Evidentemente, la retórica que acompaña a estos cambios, por parte de quien los patrocina, es que suponen una ganancia en la libertad social y personal, una recuperación en la capacidad de iniciativa y de actuación libre. Pero ¿es esto así? O, planteando la pregunta en una clara posición de sospecha ¿Los estados ceden (devuelven) poder?

Pondré un ejemplo. Como parte del proceso de integración europea en el que estamos inmersos, a partir del año 1995 se pusieron en vigor los llamados Acuerdos de Schengen sobre libre circulación de personas y mercancías entre los países de la Unión Europea. Según esos acuerdos, desparecían los controles aduaneros entre los países afectados. Antes de ese año, en el aeropuerto de Barcelona era posible circular libremente por la primera planta (una planta llena de tiendas y cafeterías), salvo unos metros antes de las puertas de embarque en donde se producía el control de pasajeros para su acceso al avión. A los pocos meses de aplicación del acuerdo, acompañaba al aeropuerto a un familiar que partía de viaje y tras los trámites de billetes, me disponía a acompañarlo a la primera planta para esperar junto a él, tomando un café, hasta la hora de partir. Cuál no sería mi sorpresa cuando me encontré con que no se nos permitía el paso a dicha planta a quienes no tuviéramos la correspondiente tarjeta de embarque. Incómodo, pregunté por las razones para ello, y me respondieron que era por la aplicación de los acuerdos de Schengen. Efectivamente, toda la primera planta pasó de ser un lugar de "libre circulación" a un lugar de "acceso restringido a la libre circulación de pasajeros". Al desaparacer la clara distinción entre vuelos nacionales e internacionales, todo el espacio del aeropuerto pasa a ser un lugar de acceso restringido. Desaparecen las fronteras, pero los nuevos espacios fronterizos parecen hacerse a la vez que más flexibles, más amplios.

¿Devuelve poder el estado? Quizás debamos ser más cautelosos y en primera instancia reconocer que lo que ocurre es que los controles cambian en sus formas, se redefinen, se reterritorializan: cambian en los lugares y en las formas. Para mí, este ejemplo del aeropuerto es una metáfora respecto a lo que significan los nuevos cambios sobre la autonomía de las instituciones auspiciada desde el estado. Y es un ejemplo que creo que nos advierte que debemos mirar a los cambios políticos como transformaciones que generan nuevos efectos de control y nuevas redistribuciones de la manera en que se ejerce el poder (Popkewitz, 1994). Deberíamos quizás entender que muchos de estos nuevos procesos que estamos viviendo suponen una transformación sobre los controles, los lugares y formas en que se ejerecen, cambios que aparentemente son una entrega de poderes y capacidades de decisión a sectores más cercanos a los afectados, pero que luego resultan ser una nueva manera de situar la relación entre quienes ejerecen el control y quienes creen haberlo conquistado. La situación escolar no está al margen de estas paradojas: qué está cambiando, en qué dirección, qué parece que ganamos y qué estamos perdiendo. Y lo preocupante de esto es si no estaremos ante una mayor apariencia de libertad que puede impedir una conciencia de nuestra falta de libertad y de los nuevos sutiles controles, quizás más débiles en apariencia, pero quizás también más extensos, a los que estamos sometidos.

Podría pensarse que analizar los actuales cambios desde la teoría de la sospecha es excesivamente tendencioso, que deja poco espacio para encontrar nuevos significados y posibilidades de acción. Puede ser. Pero entonces, lo menos que se puede decir al respecto es que todos estos cambios no tienen una lectura fácil. No son simplemente lo que prometen. Dependen de un contexto más amplio en el que significan una cosa u otra, están afectados por otros cambios hacia los que no se nos dirige la atención, y también por las mentalidades que, como producto de toda la red de cambios, vamos desarrollando, y bajo las que les damos un significado, o desarrollamos una actuación. El discurso y la política de la autonomía de las escuelas, para ser entendidos en su significado y alcance real, deben ser situados en el contexto de cambios recientes en las políticas sociales y en la política de las organizaciones. Autonomía, descentralización y devolución de poder a la sociedad y a los profesionales constituyen tendencias que deben ser interpretadas en un marco más amplio para ver lo que significan realmente, hacia dónde dirigen nuestra imaginación y nuestras energías o qué es lo que tienen de ambiguo o de paradójico.

Una de las cosas que habrá que hacer es prestarle atención a la retórica completa de estos nuevos cambios, al conjunto de sistemas ideológicos que se introducen juntamente con la autonomía profesional e institucional, y con los nuevos sistemas de producción. Por ejemplo, no creo que podamos desligar el sentido de los nuevos modos de producción y consumo (o de provisión de servicios públicos y de su elección por los usuarios) de una cultura de competitividad, rentabilidad y eficacia y de una ética que combina la identificación con los objetivos de la empresa (o del estado) con el autointerés de los particulares como modos de motivación (Ball, 1997). (Note 5)

Uno de los aspectos paradójicos del discurso y las propuestas sobre la autonomía se encuentra en que parecen ser a la vez nuevos sistemas de regulación del trabajo y de las instituciones y nuevos espacios de acción más libre y creativa. Se constituyen como procedimientos para la identificación con los objetivos impuestos a la vez que formas de aumentar la satisfacción con el trabajo. Una de las paradojas que tendremos que analizar es el hecho de que el discurso sobre la autonomía del profesorado y de los centros educativos esté instalado no en estos agentes e instituciones, sino en la academia y en la administración. Aunque la forma en que he titulado mi exposición (¿Autonomía por decreto?) puede dar a entender que me refiero a lo paradójico que resulta el hecho de que la administración "imponga" la autonomía a los centros, no quisiera que esto sirviera de distracción a la otra paradoja importante: somos los académicos, investigadores y profesores universitarios quienes estamos a vueltas con esta cuestión, igual que con la del profesionalismo del profesorado.

La división social del trabajo en educación convierte en normal algo que debiera resultarnos chocante: se habla sobre la autonomía de los enseñantes desde un lugar que parece fuera de los propios enseñantes. Se habla de la autonomía como una aspiración que se tiene para ellos, pero que no parece nacer de ellos. ¿A qué obedece dicho interés? ¿Desde qué interés hablamos los académicos en defensa de la autonomía del profesorado o de los centros escolares? ¿Desde dónde hablo yo?

Los discursos universitarios, al dirigir nuestra mirada sobre otro lugar distinto al propio desde el que se habla y al proyectar aspiraciones nobles en apariencia, pueden estar ocultando los ventajas académicas que se procuran desde dichos discursos. Diversos autores (Labaree, 1992; Popkewitz, 1994; Hargreaves y Goodson, 1996) han señalado los beneficios para el mundo universitario de la pedagogía que supone la preocupación por el profesionalismo de la enseñanza, ya que supone "elevar el estatus de quienes tienen que aumentar el estatus" del profesorado (Labaree, cit. por Popkewitz, 1994, p. 206). Pero también, la defensa de la profesionalidad, los nuevos planes de formación del profesorado no son ajenos a otros fenómenos asociados, pero a los que parece que no solemos dirigir nuestra atención todo lo que debiéramos. Es el caso, por ejemplo, de la aparición y defensa interesada de nuevos profesionales de la educación, o de la progresiva tecnologización de la práctica pedagógica. (Note 6) Paradójicamente, la autonomía del profesorado y de los centros educativos (y en general, toda la defensa del profesionalismo docente y de las nuevas reformas educativas) está dando lugar al desarrollo de nuevas figuras de "experto" que crean nuevas dependencias del profesorado respecto del conocimiento académico, conservando así el estatus superior de quien "desarrolla" la autonomía del profesorado (Contreras, 1996; Gentili, 1997; Pérez de Lara, 1998).

Este estado de cosas no es ajeno tampoco a las condiciones en que tienen que trabajar las universidades, también sometidas a las mismas nuevas formas de producción y organización del trabajo. La productividad, la evaluación por objetivos, la empresarialización, la competitividad y el autointerés, la comercialización, en definitiva del trabajo intelectual, son parte de las condiciones de trabajo también en la universidad, y en las que hay que reinterpretar el sentido de la actual "autonomía intelectual" del trabajo académico (Alonso, 1997; Neave, 1998). De una forma cada vez más evidente, la financiación del trabajo académico depende de la dirección que éste tome. La propia universidad pública sólo consigue fondos si éstos se dirigen a los objetivos de los patrocinadores, sean éstos públicos o privados. Y como ha dejado claro Lundgren (1989), el desarrollo de la investigación educativa está claramente determinado por los objetivos políticos de la Administración como financiadora fundamental de la misma.

En este contexto, hay que entender que los intelectuales, desde sus posiciones ideológicas, o desde sus intereses de clase, están siendo la vía de expresión y legitimación de políticas, así como los expertos técnicos en su desarrollo operativo y en su justificación ideológica (arropada en recursos de legitimación de los que dispone el mundo académico). El lenguaje académico, atrapado en ideologías de progreso (Popkewitz, 1994), tiende a hablar de grandes conceptos, como cambio, innovación, calidad, autonomía, como términos que remiten a significados positivos y homogéneos que nunca se aclaran y que se desvisten de lo que les daría su auténtica significación: qué cambio, qué innovación, qué autonomía, etc. Al presentarse el profesionalismo y la autonomía como valores en sí, se convierten en ideas fácilmente instrumentalizables por otros intereses, ya que no se analiza el complejo semántico, ideológico, retórico y político en el que se insertan, utilizan y operativizan las propuestas. (Note 7)

Precisamente por esto, entiendo que la función intelectual fundamental que debe ejercer el trabajo académico, si quiere evitar la tentación y la trampa del uso de las retóricas interesadas, dirigidas a fines que no son los que aparentemente anuncian con su lenguaje, es la de desvelar las tramas discursivas en las que se sitúan estos conceptos y propuestas, poniendo en relación dichas retóricas con el resto de la red ideológica y con los significados operativos y las consecuencias prácticas de las políticas generales en las que se insertan. Se trata de "desnaturalizar" y hacer visible lo que quiere ser presentado como la tendencia inevitable de los hechos y el significado incuestionable de los valores en los que se justifican. Y en esa desnaturalización, mostrar las formas en que las políticas, las prácticas, los discursos en que se sostienen nunca son, aunque se presenten así, soluciones definitivas a los problemas. Las nuevas propuestas lo que hacen es cambiar la forma en que se manifiestan las circunstancias de la vida, presentando, en el mejor de los casos, las paradojas por las que lo que parecía un valor indiscutible puede venir acompañdo de una pérdida insospechada. Atender a lo que de paradójico pueden tener hechos como las nuevas formas de organización de los servicios, o aspiraciones humanas y profesionales, como la de la autonomía, puede ponernos sobre aviso, a la vez que mostramos la riqueza potencial de lo que puede ser de muchas maneras y significar más de una cosa. Sin embargo, es conveniente una nota de precaución.

Recientemente, Hargreaves (1997) en una interesante autocrítica, ha puesto en cuestión el lenguaje complaciente de las paradojas (que él mismo había utilizado para describir ciertas tendencias actuales en el cambio escolar) como una forma atractiva y casi lúdica de mirar a la realidad. Parece que la paradoja nos sitúa en el campo de la curiosidad que incita nuestra mente en la comprensión de tales fenómenos. Pero también es cierto que lo que tal aproximación puede ocultar (como viene a decir el propio Hargreaves) es que, en primer lugar lo que presentamos (visto desde el avión, mientras revisamos una conferencia sobre el profesorado que se ve allí abajo) como una curiosa y atractiva paradoja, puede ser una experiencia dolorosa o frustrante para quien es víctima de ella, a diferencia de quien disfruta del juego intelectual que sugiere, o de la pasión investigadora que despierta. De otra parte, el lenguaje de la paradoja nos hace pensar en dichos fenómenos, así analizados, como si poseyeran una dialéctica estable y atractiva, perdiendo de vista su origen y su naturaleza interesadamente provocada. Las supuestas paradojas no son muchas veces más que juegos del poder para, haciendo aparentar una cosa, dar lugar a otra. Leerlo como paradoja puede ser un error de apreciación de lo que quizás sería mejor presentar como trampas, como intereses aparentemente contradictorios, pero unidos por una lógica oculta que conviene desvelar. Afianzar la visión de ciertos fenómenos como paradójicos puede ser un placer intelectual, pero puede impedir a quienes son víctimas de ellos a entender por qué pasan determinadas cosas que parecen afirmar una cosa y su contraria, sirviendo sólo para paralizarles en su capacidad de acción creativa.

Parte de nuestro trabajo es, por consiguiente, denunciar la paradoja allá donde es provocada ficticiamente; desentrañarla allá donde tenemos que aprender a convivir con ella; captar dónde y cuando nos ha conducido a pensar de otra manera, o dónde nos está impidiendo pensar con claridad. Lo cual no tiene por qué ser un impedimento para que a la vez aprendamos a convivir con la complejidad, sabiendo pensar complejamente, esto es, no simplificando los fenómenos complejos (Arnaus, en prensa). Como profesores universitarios, tal tarea no es sino la auténtica tarea educativa, dirigida a favorecer la consciencia de la gente sobre sus propias concepciones y condiciones de vida. En el fondo, es tanto como aceptar, para darle un sentido de posibilidad, la paradoja y tensión de la división social del trabajo en educación: utilizar la posición laboral de intelectual para crear nuevas posibilidades de pensamiento y acción, pero de forma tal que no ahonden la desautorización del profesorado (re)produciendo su dependencia intelectual. Pero enfrentado a mi propia tarea educativa, yo también me descubro como un enseñante, sometido a los mismos problemas y a las mismas retóricas sobre la autonomía en un contexto de creciente comercialización del sentido de la formación universitaria. En cuanto que enseñante, no siempre tengo que mirar a otros docentes para tratar de entender y procurar dar sentido a la complejidad de un trabajo en el que además, por dignidad humana, uno aspira a una mayor autonomía personal y profesional.

2. Los cambios en la institución escolar: del cambio educativo a las reformas políticas

¿De dónde procede toda la preocupación por la autonomía? ¿En qué contexto de cambios generales sobre el mundo escolar se está introduciendo ahora esta idea y qué es lo que tiene de paradójico: de ambivalencias y de ambigüedades, pero también de contradicciones y de trampas, y quizás también de posibilidades creativas? Un elemento que me parece de especial importancia para entender parte de las paradojas sobre la autonomía del profesorado y de los centros escolares es que ésta se ha convertido en un referente común en la teoría de la innovación y del cambio educativo. A lo largo de los últimos años, se ha producido en este campo de estudio una abundante producción, así como una sorprendente evolución en perspectivas, teorías y enfoques. Ello es producto, tanto de las insuficiencias percibidas en la forma de interpretar los procesos de cambio, como también de las modificaciones que se vienen produciendo en estos últimos años en la forma de promoverlos.

Por ejemplo, Fullan (1998) acaba de hacer una revisión de su evolución personal como estudioso del cambio educativo en estos últimos 25 años. En él distingue tres etapas por las que ha pasado su forma de entender dónde se encuentran las claves del mismo, etapas que reflejan no sólo su personal visión, sino también los hitos de la evolución en el campo: en una primera época, fue la idea de *implementación* la que dominó su teoría del cambio. Al destacar el papel fundamental de quien tenia que "adoptar" y realizar las innovaciones, su análisis permitía criticar los modelos imperantes hasta entonces de innovación, y el fracaso de los mismos para provocar cambios significativos. Eran innovaciones pensadas de arriba a abajo, elaboradas fuera de la escuela y luego transmitidas dentro, sin tener en cuenta los factores personales y contextuales que afectan a la aplicación de una innovación, y desconsiderando también el papel activo que podían tener los enseñantes en su desarrollo. Pero como el mismo Fullan reconoce, la recuperación del papel de los enseñantes en esta primera etapa se seguía haciendo bajo una perspectiva lineal y dependiente, como implementadores activos de innovaciones ya pensadas.

La segunda época que señala Fullan tiene como preocupación central el *significado* del cambio, esto es, la pretensión de entender fenomenológicamente la diferencia entre cómo se pretende el cambio y cómo lo experimenta realmente la gente, en cuanto que factor clave para comprender el fracaso de la mayoría de las reformas sociales. Para ello, necesitaba entender los diversos factores, macroorganizativos, institucionales y personales. Una creciente comprensión de la forma en que estos diversos factores estaban presentes en los procesos de cambio, le permitió darse cuenta de que el punto de partida del cambio educativo no debía situarse tanto en las innovaciones como en los individuos y en las instituciones.

La tercera época, la actual, la denomina este autor la de la *capacidad de cambio*; en esta época sitúa las fuerzas fundamentales del cambio en el individuo y en el grupo, dentro de una organización y una sociedad con capacidad de aprendizaje: considera que el núcleo de la reforma educativa lo constituyen el desarrollo individual y el institucional. La cultura institucional de la escuela debe transformarse, de manera que admita formas más variadas e interactivas de profesionalidad; pero son los enseñantes concretos quienes sostienen y por tanto pueden transformar la cultura del centro.

> "No podemos depender de, o esperar a que el sistema cambie... Debemos desarrollar nuestras propias capacidades para aprender permanentemente, y no dejar que las vicisitudes del cambio nos depriman." (Fullan, 1998, pág. 224).

Otra especialista en el cambio educativo, Rudduck (1994), ha presentado los diferentes estilos de reforma que se han sucedido en los últimos tiempos en Inglaterra, no ya como tendencias en el estudio, sino como formas de innovación que se han intentado llevar a cabo. Según esta autora, una secuencia semejante puede reconocerse en la mayoría de los países: una primera fase, de *reforma curricular*, propiamente dicha, en la que se pusieron en marcha grandes proyectos curriculares. El problema de este tipo de experiencias de innovación fue que si bien pudieron estar concebidas como proyectos coherentes, dejaban al profesorado en una posición pasiva respecto al mismo, como aplicadores, sin tener en cuenta las condiciones específicas en las que realizaban su trabajo; ademas, al ser proyectos normalmente referidos a áreas y materias particulares, sólo afectaban a algunos profesores dentro de una misma escuela, lo que normalmente provocaba conflictos dentro de los centros, los cuales, en ocasiones, podían encontrarse desarrollando simultáneamente diferentes experiencias de innovación con lógicas distintas. Una segunda fase, que pretendía evitar los problemas anteriores fue la del *desarrollo curricular basado en las escuelas*. En esta ocasión, los proyectos pretendían ganar en relevancia local, pensándolos como procesos internos a los centros, atendiendo a sus propias necesidades, pero eran procesos que requerían el despliegue de una gran energía y en donde no siempre se disponía de suficientes recursos para dotar al proceso de coherencia interna o para tratar los conflictos que se pudieran generar. La tercera fase que reseña Rudduck es la del *desarrollo de la escuela como totalidad*. En este caso, la preocupación se centra, más que en el desarrollo de programas, en la construcción de un sentido colectivo, dentro del propio centro, tomando algún propósito compartido como eje de desarrollo y aprendizaje de la escuela.

En definitiva, la consideración del profesorado y de las instituciones escolares como mediadores activos en los procesos de cambio, de los centros como unidades y como culturas totales, del cambio como un proceso continuo de aprendizaje, y de los enseñantes como fuerzas individuales que pueden dirigir sus energías en cuanto que agentes activos, se han convertido en argumentos de peso para el apoyo de reformas en las que se reconozca la autonomía del profesorado y de los centros escolares. Pero sería ilusorio y reduccionista pensar que los cambios de concepción que se producen respecto a la institución escolar o al papel de los enseñantes sólo vienen determinadas por las conclusiones de los especialistas en el cambio educativo. Suponer que el auge en la preocupación por la autonomía depende de factores relativos al éxito o fracaso de experiencias de innovación es creer que las transformaciones en la educación dependen sólo del cambio planificado sobre ella o de las conclusiones de los especialistas en cambio planificado. Por ello, si queremos ampliar nuestra comprensión del fenómeno de la autonomía debemos dirigir nuestra mirada hacia factores sociológicos y políticos.

Dubet y Martuccelli (1998), refiriéndose a Francia, han analizado los cambios que se han producido en los últimos tiempos en la institución escolar, desde un sistema estratificado y regulado, a otro de masas y desregulado. Es importante resaltar que de lo que hablan estos autores no es sólo de lo que han sido los cambios materiales, estructurales y funcionales, en el sistema escolar, sino sobre todo en lo que podríamos llamar la "imagen social" de la escuela, esto es, en la forma en que ésta ha sido percibida. Evidentemente, ambos niveles se alimentan mutuamente, pero no de una forma mecánica, lo que explica que determinadas percepciones o imágenes sociales puedan no corresponder con lo que la escuela es, pero afectan a la forma en que es vivida, y por tanto, a la forma en que la gente, sus diversos públicos, se relacionan con ella. (Note 8)

En opinión de estos autores, mientras la escuela se constituyó como una institución elitista, mantuvo también la imagen de una institución internamente justa, en la que quienes "valían" eran elegidos para continuar estudios. En todo caso, era la sociedad quien era injusta al impedir el acceso a la escolaridad a grandes capas de la población. La lucha durante mucho tiempo fue por el acceso a la escuela, que parecía la única institución justa en una sociedad injusta. Suponía el acceso a una cultura, ciertamente elitista, alejada de lo social y de la utilidad inmediata o laboral, separada de la cultura no escolar. Una institución en donde sus actores se incorporaban a los roles establecidos, indiferentes a las necesidades y características de las personalidades específicas, dirigidos todos a los fines unificados de la escuela. Eso la convertía en homogénea en cuanto a contenidos y procesos pedagógicos (el mismo programa, las mismas reglas impersonales), con una regulación mecánica en donde lo importante es cada aula con su enseñante y su grupos de estudiantes. En todo caso, las divisiones son producto de la estratificación social, en donde los públicos están separados y jerarquizados, cada uno con sus formas escolares y sus objetivos diferenciados (escuela pública y privada, escuela primaria y secundaria, bachillerato y formación profesional, etc.). Era, pues, una institución con un fuerte consenso a su alrededor, con una fuerte definición de roles, normas y valores, con un claro sentido de las expectativas y del futuro que ofrecía.

Sin embargo, la extensión de la escolarización a toda la sociedad y la ampliación de la obligatoriedad, juntamente con las críticas que recibió este sistema escolar, han ido cambiando no sólo las formas escolares sino también la forma en que es percibido. Un primer efecto de esta masificación escolar es la modificación de los sistemas selectivos; ya no es suficiente con la clasificación social previa a los estudios, sino que ahora se produce en el flujo de los recorridos escolares; si estudiar ha dejado de ser un modo de distinción, las nuevas distinciones buscan producirse dentro de la escolaridad, y no ya entre estudiar y no estudiar; si ya no hay un control estricto en el acceso a la secundaria, ahora es el proceso que se sigue dentro de ella quien cumple las funciones selectivas: los resultados escolares, las calificaciones, tienen ahora una especial importancia en la orientación de las posibilidades, de forma que es en el interior de la escuela en donde empieza a producirse la diversificación de las oportunidades.

Un segundo efecto, asociado al anterior, es el de la diversificación continua dentro del sistema; continuas ramificaciones jerarquizadas dentro de la escuela vienen a sostener el sistema de clasificación interna; si ya no se trata de la división entre quienes van al bachillerato y quienes no van, ahora la diferencia la proporciona el tipo de bachillerato. La diversificación da lugar a la competencia interna del sistema (competencia entre itinerarios con distinto valor, oportunidades y consecuencias que se recomiendan en función de los logros y las expectativas, competencia entre disciplinas que se atribuyen diferente prestigio y valor escolar, entre centros, que procuran atraerse a los mejores estudiantes para mantener una reputación atractiva). Esta competencia interna del sistema lleva asociada una competencia por parte de los usuarios, que se manifiesta en su capacidad estratégica en la utilización del sistema en su propio beneficio; ya no estamos ante una incorporación a la institución escolar, a la espera de un proceso de formación, lo cual significaba el reconocimiento de una autoridad, la asunción de un rol y la aceptación de la misión que la institución prometía. Cada vez más el alumnado juega sus cartas en la oferta de la escuela, buscando una rentabilización de sus acciones, y las familias procuran orientarse en el sistema de selección. Y de la misma forma, es la escuela la que ahora es percibida como institución injusta al acumular mecanismos selectivos en su interior.

Producto en parte de las críticas recibidas por la escuela como institución alejada de los intereses y necesidades de la infancia y la adolescencia a la que atiende, y también de las críticas por la falta de utilidad y de rentabilidad en la formación para el trabajo, el sistema educativo se ha visto sometido a una doble demanda de abrir la escuela a la infancia y de abrirla a la economía, mientras que, en oposición, algunos sectores pretenden mantener la tradición anterior de contenidos y relaciones pedagógicas. La concurrencia de diversas finalidades para la educación ha dado lugar a la aparición de nuevos conflictos, así como a la convivencia de propósitos no siempre compatibles o fácilmente conjugables. Por ejemplo, la atención a la diversidad del alumnado se ha pretendido juntamente con la igualdad de resultados; el respeto a las personalidades variadas del alumnado supone también la incorporación de nuevos conflictos, generados por la propia diversidad, a la vez que las discrepancias entre culturas profesionales, respecto a las fórmulas organizativas y pedagógicas para ello. De igual modo, la diversificación de ofertas e itinerarios para atender a las demandas de eficacia social y a la rentabilidad ha chocado con la resistencia de la cultura escolar académica, alejada de la social y de la idea de utilidad (a veces tanto por razones progresistas- críticas como por motivaciones conservadoras-elitistas). El resultado de todos estos cambios es que la escuela ha dejado de ser una institución de consenso para pasar a ser una institución diversificada, con una crisis difusa ante la diversidad de finalidades demandadas poco conciliables y en donde

> "la unidad de juicio se ha roto. Los profesores, la administración y los padres forman... un cuadro dramático en el cual cada uno se percibe como hostil a todos. Los equilibrios precarios reemplazan al orden "natural" de la vida escolar." (Dubet y Martuccelli, 1998, pág. 58).

La organización pedagógica se ha desestabilizado con el derrumbe de las antiguas formas pedagógicas, la "caída" del nivel, la pérdida de prestigio del profesorado, la competencia de la cultura escolar con las atractivas y poderosas culturas de masas, la presencia en las aulas de alumnos conflictivos. La escuela ya no es más una estructura institucional bien definida y estable en sus normas y contenidos, con reglas que puedan emanar del centro administrativo y ser cumplidas homogéneamente en todos los centros. Ya no hay soluciones generales, sino que deben ser hechas a medida. Ahora cada centro y cada enseñante deben elaborar sus propias políticas.

> "El establecimiento escolar ya no es solamente el eslabón de una cadena burocrática: está obligado a construir una "política" ajustando las prácticas de los docentes y de los alumnos y dándoles un cierto dominio de su entorno. La escuela es una organización de fronteras flotantes, de objetivos cada vez redefinidos, de relaciones cada vez reconstruidas; ya no es reductible a la forma burocrática general que la encierra." (Dubet y Martuccelli, 1998, pág. 60).

Como podemos ver, estos cambios sitúan el tema de la autonomía del profesorado y de los centros en otro orden de problemas bien diferente del señalado por Fullan, como capacidad de aprendizaje y de innovación, como fuerzas de cambio. Si la autonomía puede ser realmente el reconocimiento de que las energías del cambio nacen de la capacidad de las escuelas, al margen del sistema, también debemos entender que hay algo en su origen que es consecuencia de profundas alteraciones en el sentido de la escolaridad y en los problemas que acarrea. Los estados, las organizaciones burocráticas y administrativas han perdido legitimidad y capacidad para administrar desde el centro los conflictos y los traspasan a la propia capacidad y responsabilidad local de los establecimientos educativos particulares (Weiler, 1990).

Esta situación a la que se ha visto abocado el sistema educativo, junto con los nuevos modos de entender los servicios públicos y la regulación del trabajo, de base post-fordista, alentados por una nueva ideología política centrada en la productividad, y amparada por la cobertura intelectual de las nuevas estrategias del cambio educativo, constituye la confluencia de intereses y circunstancias que ha dado lugar a las políticas de la autonomía de centros escolares. En resumen, podríamos destacar los siguientes rasgos de esta nueva política.

1. La idea de un curriculum común y una escuela común, como un lugar filtrado y selectivo, dirigido por la idea de la meritocracia de los individuos, ha dejado de funcionar. La variedad de alumnado sociológicamente distribuido y desigual ha convertido en principio político lo que era una práctica real: el curriculum y la escuela deben adaptarse al alumnado. La adaptación del curriculum es el reconocimiento de que vivimos en una sociedad plural, con características y necesidades diferentes, pero también

con aspiraciones e intereses variados. Pero lo que supone la nueva política curricular no es sólo un reconocimiento pasivo de este hecho, sino también una nueva exigencia: la sociedad y el alumnado deben ser reconocidos como diversos y plurales, y es función de centros y enseñantes considerar la relación entre tal diversidad y la forma en que concretan el curriculum. El acento ya no se pone en la idea de un curriculum igual para todos, sino en la de un curriculum para la diversidad.

2. El mensaje de un curriculum adaptado a la diversidad social no es sólo para los enseñantes, sino también para la sociedad en general: vivimos en una sociedad variada y la escuela (o mejor dicho, las escuelas, esto es, cada una de ellas) deben adaptarse a este hecho. Si la sociedad es variada, las escuelas deben serlo también. No sólo debe mirar la escuela a lo variado de la sociedad, sino que también la sociedad debe atender a lo variado de las escuelas. Es así como el derecho o la posibilidad de elección se convierte ahora en una obligación.
3. La idea de un curriculum descentralizado y para la diversidad no significa, sin embargo que el estado abandone el control de un curriculum nacional ni los procesos de regulación acerca de la forma en que deben desarrollarse localmente las reformas. Precisamente una de las características de las nuevas políticas de los servicios públicos está en la fijación por el estado de los objetivos de productividad. La legitimidad del estado se ve doblemente reforzada en la medida en que estas reformas (esto es, la fijación del curriculum nacional, su contenido y su justificación) se hacen más aceptables, al abandonar la responsabilidad sobre conflictos sociales asociados a la educación (que ya no podrían achacarse a la política central, sino a los propios centros), y en la medida en que este abandono de responsabilidad consigue presentarse satisfactoriamente como una devolución a la sociedad y a los enseñantes de espacios de decisión, manteniendo el control central sobre aspectos fundamentales del curriculum y de la gestión educativa.
4. La autonomía de centros parece cobrar aquí su significado expreso como forma de organizar la aspiración a que los distintos centros educativos asuman responsabilidades en la adecuación de su actuación a las características y necesidades sociales y personales del alumnado al que atenderán. Se espera que los centros no sean ahora sólo el espacio material en el que ocurre la enseñanza, sino también el lugar orgánico en el que se configura "autónomamente" un plan específico para gestionar el encuentro entre un curriculum nacional y una realidad social concreta. La "devolución" de competencias a los centros públicos es la forma de encomendarles el desarrollo y la concreción de los planes administrativos, de forma que puedan ser adecuados a la variedad social. Pero también, en buena lógica con las nuevas tendencias en innovación educativa, se espera que los centros funcionen también como entidades con identidad propia, que generen un sentido de pertenencia entre el profesorado. De esta manera, al sentirse más implicados en la concreción administrativa de la

reforma, asumiendo responsabilidades respecto a la misma, se espera que se involucren más en sus principios y en sus prácticas.

5. Que cada centro asuma "autónomamente" la reponsabilidad de su propio proyecto educativo tiene su traducción práctica en que ante quien asumen dicha responsabilidad es ante la "sociedad", y por ella suele entenderse en este caso las familias concretas que acuden a los centros escolares concretos. La sociedad, particularizada en las familias singulares con hijos en edad escolar, asume las responsabilidades "devueltas" por el estado, asumiendo la obligación de exigir a los centros una educación de calidad (esto es, aquella educación que mejor se adecue a sus valoraciones y expectativas). La forma en que se está entendiendo este principio de participación de las familias en la educación es sobre todo animando y facilitando la elección de centros. Este factor de elección se supone que animará a la comunidad educativa, de manera que cada escuela e instituto estará pendiente de su valoración pública y la "sociedad", pendiente de los mejores centros. De esta forma, como decía antes, la devolución de responsabilidades es entendida como la entrega a los actores concretos (las escuelas específicas y las familias afectadas en cada una de ellas) de la responsabilidad de los efectos de sus decisiones aisladas. Efectos que en ocasiones sólo pueden ser entendidos en su dimensión sociológica, cultural y política, y no sólo en la dimensión particular en las que estas decisiones se toman.
6. Esta importancia que adquiere el centro educativo en el contexto de las nuevas reformas ha hecho que se haya puesto un especial interés en la conducción de las mismas, en los procedimientos mediante los cuales debe desarrollarse esta nueva forma de gestión. Y para ello se elaboran nuevas pautas o reglamentaciones sobre las formas en que debe trabajar un centro para desarrollar su proyectos y para mejorar institucionalmente. Son pautas que se refieren a las formas de trabajo en equipo, a los procedimientos para la elaboración y desarrollo de proyectos educativos y a la manera de implicar al profesorado en este tipo de tareas. Uno de los factores fundamentales que se descubre en este esfuerzo por convertir a los centros en instituciones más dinámicas y preocupadas por su mejora como colectivo profesional, es el del liderazgo, esto es, la importancia de figuras institucionales concretas que asuman el papel de dinamizadores del equipo docente. Reforzar el liderazgo, e incluso profesionalizarlo (instituyendo figuras específicas con una formación también específica, o bien dignificando el cargo, o concediéndole mayor autoridad institucional a directores, jefes de estudios o jefes de departamentos), es una de las preocupaciones de estos planes administrativos para el desarrollo organizativo y pedagógico de los centros escolares.
7. Expertos y administradores insisten en la importancia de que los centros se autoconsideren como unidades que se gestionan a sí mismas, sensibles a su contexto, tratando de atender a sus demandas y en continuo desarrollo profesional e institucional.

> Los enseñantes harán suya la reforma si hacen suyo el curriculum y si se comprometen con su centro educativo, confiriéndole un carácter propio y singular. La calidad de la educación depende de la calidad de los centros, y ésta a su vez depende de que el profesorado se comprometa con los mismos, de que trabaje en colaboración con sus compañeras y compañeros para su permanente mejora, atendiendo a las necesidades del contexto y dando respuesta a las demandas. El trabajo profesional del profesorado ya no se circunscribe al aula y al alumnado concreto al que se enseña, sino que ahora abarca la preocupación por el centro como unidad educativa. La responsabilidad de los centros públicos (de su evolución, de su calidad y de su futuro) ya no es solamente del estado, sino que también lo es del equipo docente que trabaja en ellos. Ahora, la valoración que pudiera hacerse sobre la calidad de una escuela o de un instituto no podrá disociarse de la valoración sobre la calidad de su profesorado.

3. Análisis de las actuales políticas de autonomía de centros (Note 9)

¿Supone este cambio de perspectiva realmente que los enseñantes pasan a ser sujetos autónomos en la definición del cambio educativo, que se convierten en generadores de nuevas prácticas docentes y en los impulsores de nuevas direcciones para la educación escolar? ¿De qué manera? ¿En qué dirección?

Mi hipótesis es que la "autonomía por decreto" no es sino el reflejo de formas más sutiles de control mezcladas con formas de responsabilización del profesorado en la elaboración de servicios adaptados a la clientela, junto con una pretensión, en muchas ocasiones honesta, pero probablemente desenfocada y equivocada, de racionalización de una práctica que ha perdido su potencial transformador al crear modelos universales de cambio que han perdido su significado político. Porque la auténtica autonomía profesional no existe al margen de la definición del sentido político de la misma, y este sentido político está relacionado con la autonomía social en la que busca desarrollarse la profesional y en la que busca su significado potencial.

Deberíamos poder calibrar cuándo el cambio educativo que se pretende está propiciando valores educativos y sociales dirigidos a favorecer que las personas, los colectivos y las comunidades sociales se hagan más autónomas, más dueñas de sus propias vidas y con mayores posibilidades de dirigir su futuro (en un contexto de cohesión social y solidaridad en la diversidad), y cuándo estamos ante pretensiones de control (y por tanto de dependencia profesional y social respecto de propósitos que no son los propios), sólo que ahora dirigido hacia la capacidad de autocontrol.

¿Cuál es el análisis que cabe hacer sobre el sentido de estos cambios, sobre su significación, su dirección y sus (posibles) consecuencias?

- A) Competitividad y oferta-demanda: movimientos particulares, no diálogo social Cada vez se entiende más la escuela como un bien de consumo para el ciudadano en el que éste debe manifestar su participación en la oferta de bienes escolares, eligiendo unos y rechazando otros. La escuela es más democrática —se piensa— si tenemos más capacidad de elegir la que nos interesa. Es decir, si las escuelas son más diferentes entre ellas. Esto conduce a las escuelas a la competitividad en la que el mercado de oferta y demanda debe ajustarse.

  Es evidente que este planteamiento de la competitividad y del ajuste entre la oferta y la demanda tiene su perversión. En primer lugar, ni la oferta es igual para todos, ni la capacidad de opción es equivalente entre todos, dadas las diferencias económicas, sociales y culturales de partida entre los usuarios. Implicar a las familias en la elección es hacerlas responsables de sus recursos personales de elección (ya sean económicos, culturales, de dedicación, etc.). Pero los recursos ya están desigualmente distribuidos. En segundo lugar, la competitividad como motivación de la sociedad no es ni mucho menos neutra. El criterio a partir del cual se compite no es libremente decidido por "la sociedad", puesto que viene decidido por la capacidad de rendimiento en los términos que fija el curriculum oficial (Hatcher, 1994). En efecto, el curriculum no sólo actúa como modo de control sobre los objetivos de los centros, sino que además permite hacer evaluaciones comparables entre los distintos centros, y al hacerlo así, impulsa a éstos a intentar elevar el rendimiento según viene definido por dichos objetivos, y a las familias a asimilar que la calidad de las escuelas depende del éxito en el rendimiento. Así se evita que puedan proliferar otras filosofías y prácticas educativas que pudieran guiarse por otras pretensiones, por otros objetivos y contenidos que no sean los que ya se han convertido en patrón de competitividad.

  Lo que refleja todo esto es efectivamente un modelo de ajuste y demanda, pero no un modelo de diálogo social en la definición de la escolaridad. Escuelas y usuarios resultan estar haciendo movimientos de ajuste a partir de demandas y necesidades que no controlan, porque no actúan como colectivo que toma decisiones deliberativas y compartidas, sino como agentes aislados guiados por intereses individuales, no sociales. No participan en la definición colectiva de la educación y de sus vinculaciones con la sociedad, sino tan sólo en procesos de elección y de adecuación a partir de decisiones de las que están excluidos, porque son decisiones que toma la Administración.

- B) Descentralización: salto en el vacío La descentralización se interpreta como el salto en el vacío de las administraciones centrales a los centros particulares, de tal forma que las escuelas quedan convertidas en islas (Brennan, 1993). La autonomía escolar se plantea sobre la base del aislamiento de los centros. Esto significa, de una parte, la ausencia de estructuras sociales intermedias con capacidad de creación y actuación política, y no sólo de ejecución administrativa (Angulo, 1992). La autonomía escolar no consiste por tanto en sistemas de acción social organizada para articular lo escolar de una forma más amplia de lo que es posible desde las escuelas individuales. De otra, el aislamiento de los centros abona la posibilidad de que el desarrollo curricular y educativo en general que se espera de ellos quede reducido a lo que pueda ser generado desde los recursos aislados de cada escuela. Pero difícilmente pueden los centros aislados generar políticas y recursos adecuados para un desarrollo curricular que requiere algo más que capacidad de decisión sobre la organización y secuenciación de los contenidos, esto es, para un desarrollo curricular que necesita capacidad de acción transformadora de las condiciones de enseñanza. En el aislamiento, la mayoría de las escuelas e institutos lo único que pueden hacer es identificar el desarrollo curricular con el desarrollo de documentos: los Proyectos de Centro. Además, el aislamiento, como apariencia de autonomía, anima al profesorado a centrarse sólo en su escuela, más que a desarrollar una perspectiva profesional según la cual comprender la relación entre la enseñanza, el sistema educativo en general y las condiciones sociales y políticas de su trabajo (Angus, 1994).

- C) Paradojas control-abandono y flexibilización- regulación: Autoadministración y despolitización Una de las paradojas de las actuales políticas de reforma es que combinan simultáneamente el control de la educación por parte del estado y su abandono al mercado. Por esta razón, la supuesta autonomía refleja la contradicción entre la flexibilización y la creciente reglamentación. Pero esta contradicción es más aparente que real si entendemos que la autonomía de centros está pensada como una forma de autoadministración, que no como una forma de autogestión política. La autonomía, entendida como desarrollo administrativo queda claramente reflejado en una de los principales tareas simbólicas de la reforma: escribir el Proyecto de Centro. Es así como se explica que la autonomía de los centros se haya convertido en una complejidad burocrática y una sofisticación de las formas de gestión de los centros, más que en un proyecto político-educativo en el que implicar al alumnado y a las familias. (Note 10) Es, como refieren Lima y Afonso (1995) respecto al caso de Portugal, un concepto no democrático de autonomía, sino instrumental, como proceso de implementación. Se entiende como una técnica de gestión para articular mejor el centro con la periferia, consiguiendo una mejor realización en la ejecución de las políticas (re)centralizadas.

En la medida en que la relación con la comunidad se sustituye por la dinámica de elección, centros y enseñantes se repliegan sobre sí mismos, hacia la definición del producto educativo que ofrecen al mercado, y hacia la gestión, hacia el cumplimiento de las exigencias administrativas bajo las cuales debe realizarse tal relación de oferta y demanda, identificándose la autonomía con la autoadministración para el éxito. Esto supone que la enseñanza escolar pierde su sentido político (que significaría la conciencia de que la educación debe ser entendida como una práctica que tiene presupuestos y consecuencias sociales, por lo que debe ser objeto de participación, deliberación y acuerdo social), quedando convertida en una cuestión meramente técnica.

Toda la complejidad en el desarrollo y gestión de las actuales exigencias a las escuelas favorece esta despolitización, ya que dicha complejidad actúa como motivación para la exclusión del alumnado y de la comunidad de discusiones que han adoptado una forma muy tecnificada. A la vez, esta despolitización hace creer que el problema educativo es tan sólo uno de eficacia. Sólo que ahora esta eficacia no hace sólo referencia a los métodos de enseñanza, como ocurría hace unos años, sino al funcionamiento general de los centros educativos. El auge de la mentalidad tecnicista a que conduce la concepción gerencialista de la autonomía se basa además en una concepción ilusoria de la autonomía, porque hace creer a los enseñantes que poseen un dominio sobre su trabajo, cuando detrás se esconde la incapacidad para entender las claves políticas que lo sostienen, así como la incapacidad para enfrentarse a una de las características fundamentales de la práctica educativa: su ambigüedad, sus conflictos de valores y la permanente reinterpretación necesaria de las situaciones para dilucidar lo moralmente correcto, más que lo eficaz. A mantener esta ilusión ayuda el espejismo que provoca la retórica del profesionalismo, identificando dominio con tecnificación y haciendo creer que autonomía significa especialización y exclusividad de una práctica no compartible.

- D) Devolución sin participar en esa política: Pseudoparticipación

No deja de llamar la atención, a la vez que es muy clarificador sobre el auténtico alcance y valor de la autonomía del profesorado que representan las actuales reformas educativas, el hecho de que la supuesta autonomía concedida no sólo no nace de una demanda del profesorado (Gimeno, 1994), sino que éste ni siquiera participa en su proyecto político. Paradójicamente, al profesorado y a las escuelas se les "devuelve" algo que no han pedido y ni siquiera participan en la formulación de las políticas de devolución. Esta paradoja se convierte necesariamente en otra: se concede autonomía escolar, pero los parámetros de las políticas ya están fijados (Robertson, 1993).

Que esta aparente autonomía no significa gran capacidad de acción lo demuestra el hecho de que la capacidad de decisión e intervención sólo es "concedida" a partir del centro, siendo excluido el profesorado de todos los presupuestos, condiciones y organizaciones previas. Una vez que otros, otras instancias han decidido y planificado todo hasta las puertas de la escuela, los enseñantes actúan, pero sin capacidad de intervenir "hacia arriba" para transformar las condiciones en que se les deja "autónomos" Son sólo obedientes funcionarios que deben resolver los problemas en los que se encuentran, tal y como se los han definido y con los recursos que les han concedido. Su autonomía no es política, sino tan sólo de gestión. Como ha dicho Angus (1994), una nueva burocracia ahora disfrazada de participación. Es lo que Bottery (1992) ha llamado la pseudoparticipación, esto es, se promueve la participación del profesorado en la realización de las tareas específicas del trabajo, pero se va desanimando ésta conforme se asciende hacia la organización del mismo, y más aún en las esferas de definición de las políticas de la institución.

E) Asunción de la nueva ideología: implicación, competitividad, autointerés, control entre colegas Sin embargo, el éxito de las nuevas reformas depende de que el profesorado "sienta" que su autonomía ha aumentado, por lo que es necesaria la creación de un sentido de propiedad en un juego que no ha elegido. No se trata solamente de que realice "autónomamente" las reformas previstas, de que adaptándolas las adopte. El modelo del mercado requiere además que participe ideológicamente del principio de la competitividad, que asuma las nuevas reglas del juego, a la vez que respeta los márgenes del curriculum oficial. Según ha puesto de relieve Hatcher (1994), los mecanismos del mercado requieren la socialización del profesorado en nuevos valores: requieren pasar de una ética de servicio a un sentido de autointerés competitivo. Y en su opinión, este cambio se produce asociado a una transformación en la ideología del profesionalismo, en donde el acento pasa de situarse en la noción de autonomía, a centrarse en la competencia técnica y la competitividad en el mercado. Al fin y al cabo, el profesionalismo suele estar asociado al ejercicio liberal de la profesión, que incluye este sentido de autointerés, riesgo y ganancia.

El éxito de una política de este tipo depende por consiguiente de la forma en que la propia institución escolar y la propia cultura profesional se transforma, adoptando procesos de autocontrol, en vez de control externo. Es, como lo ha llamado Smyth (1993), un control participativo. Pero también, como lo llama Ball (1993, pág. 66), una conducción a distancia, donde "la coerción es reemplazada por el autocontrol —la apariencia de autonomía." Y es que, bajo la necesidad de coherencia y cohesión, las técnicas de gestión y los proyectos curriculares de centro pueden convertirse en una forma de control más estricto y cercano y en un impedimento del disenso de los colegas (Ball, 1997). De la misma forma que la apariencia de autonomía y la pseudoparticipación, aunque puede aumentar la capacidad de acción de los enseñantes en algunos aspectos, simultáneamente los puede hacer también más dóciles.

- F) Más responsabilidad sin poder. Meritocracia de escuelas Pero hay aún otro logro político de estas reformas: los enseñantes aumentan su responsabilidad, pero sin aumentar su poder, esto es, sin disponer de más posibilidades de transformación de las circunstancias en las que se desenvuelve su trabajo (Ball, 1993). En la medida en que más responsabilidad no significa más poder de cambiar las cosas (no sólo por los controles burocráticos, sino por la imposibilidad de cambiar la imagen de un futuro incierto, en el que la promesa de la escuela acerca de mayores oportunidades sociales queda desmentida por la presencia del paro y la marginación), mayor responsabilidad supone un mayor deterioro de la profesión y un mayor rechazo social, al recibir todas las culpas por la incapacidad de hacerse cargo de todas las misiones sociales que ahora se le encomiendan a la educación. Un efecto de esta responsabilidad sin poder es una redefinición de la meritocracia, en una época en la que los excluidos aumentan y donde la movilidad social es descendente. Al aumentar la responsabilidad hacia las escuelas se resitúa ideológicamente el principio de la meritocracia, en función de la capacidad o incapacidad de las escuelas para dar cuenta de las nuevas necesidades de la sociedad y de la juventud. Ante la falta de políticas no escolares para remediar la crisis, la política de la autonomía escolar hará ver que algunas escuelas sacan adelante a su alumnado, mientras que otras no. Así, si clásicamente la meritocracia era la justificación de que las diferencias no eran sociales, sino individuales, ahora se añade a ello que la solución de las mismas no depende sólo de los méritos propios de los individuos, sino también de la calidad de las escuelas. Serán las escuelas, las que fracasen o tengan éxito; serán ellas las responsables del deterioro social, no las políticas económicas y sociales.

4. Conclusiones: buscando un futuro

Si bien el panorama que acabo de presentar no es muy halagüeño respecto al significado que está adquiriendo la autonomía de centros y de enseñantes y las consecuencias que puede llegar a tener, no debe deducirse de aquí, como es lógico, que deba minusvalorarse la importancia de que el profesorado piense en lo que están haciendo colectivamente y trate de entender y atender a las necesidades de su alumnado. El problema no es, por supuesto, que los centros realicen sus planes y proyectos, sino más bien, en qué contexto se está definiendo esta tarea y bajo qué mentalidad hegemónica la podemos estar entendiendo y realizando.

Lo que significa esto es que, a pesar de la crítica realizada, las actuales reformas contienen en su interior elementos de paradoja. Por ejemplo, en principio puede aceptarse como una ganancia potencial el hecho de que a las instituciones escolares se les reconozca el derecho y la obligación de pensar en sí mismas en conjunto, tratando de determinar cuáles son sus propósitos educativos y cómo piensan desarrollarlos. De igual manera, hay que interpretar como un beneficio, y como la posibilidad de mayor satisfacción profesional, el reconocimiento y la legitimación de elementos de reflexión autónoma en la definición del trabajo docente. No obstante, estas supuestas ganancias pueden ser aparentes si sólo pueden generar procesos de discusión sobre los cómos de la práctica educativa, sin capacidad de decisión sobre los qué y los hacia dónde, que tienen que ser aceptados, o sin capacidad de intervención sobre las condiciones en las que transcurre la escolaridad. El éxito político de estas reformas depende precisamente de que los enseñantes acepten estas limitaciones como una ganancia. (Note 11)

Lo que es importante entender aquí es que más autonomía no significa más margen de maniobra (ya sea por parte de los centros y enseñantes, o por parte de las familias y los particulares), sino más capacidad de intervenir en las decisiones políticas por las que se ceden responsabilidades a las escuelas, así como más capacidad de intervención en las condiciones por las que tales cesiones pueden usarse para una mayor vinculación social en el desarrollo del bien común.

No puede hablarse de autonomía sin una clara conciencia del papel social y político que desempeña la escuela y cómo éste se concreta en cada caso. Esto significa no sólo una comprensión sociológica de cómo la escuela contribuye o puede contribuir a la igualdad o a la desigualdad social. Significa también una comprensión de cómo la enseñanza debe procurar dotar a todo el alumnado de recursos culturales e intelectuales socialmente equivalentes e internamente plurales. Esto es, una clara conciencia de cuáles son las bases culturales que debe proporcionar la escuela en cualquier caso y cómo esto debe ser leído y garantizado (más que diferenciado) para cada caso (Gimeno, 1994; Elliott, 1998).

En este sentido, no creo que sea defendible una autonomía de los centros si, en vez de referirse a las condiciones institucionales, sociales, políticas y de recursos para garantizar lo anterior, ésta se convierte, por el contrario, en una forma de asegurar unas "señas de identidad" o una "línea pedagógica" con la que justificar nuevas formas de control sobre el pensamiento y la práctica de unos enseñantes sobre otros, en nuevas formas de exclusión de alumnado y de somentimiento del profesorado. Los peligros de la autonomía no son sólo los de la privatización y el clientelismo, sino además todos los relativos a las presiones de los grupos de poder internos a los centros para imponer un "consenso", para excluir posiciones o prácticas pedagógicas diferentes (algo que hay que pensar que actuará más para frenar posiciones progresistas en claustros conservadores, que no al revés).

Apple (1998) se ha referido a que, en gran medida, la huida de las familias hacia modelos de elección y de mercado es producto de las propias prácticas "profesionales" de secretismo y de exclusión de la participación de la comunidad. Pero igualmente, en un contexto como el actual, en el que la autonomía y las políticas de elección les caen de arriba, sin tomar parte en las formas en que ahora se les crean nuevas reponsabilidades, la reacción docente puede ser la de la excerbación de los intereses y beneficios personales o grupales, no la de la reflexión meditada de sus responsabilidad y la reconstrucción de su compromiso con ideales educativos. La autonomía puede actuar como excusa para políticas reaccionarias en los centros si, a la ausencia de vínculos con la colectividad y formas políticas de participación en niveles más generales que los de las escuelas aisladas, se une la reacción conservadora del profesorado (amparada en la propia autonomía como derecho a sus propias decisiones) como respuesta ante políticas que les desautorizan, a la vez que les obligan a asumir riesgos y les convierten en responsables de las consecuencias de lo que no pueden determinar.

No creo que puedan desvincularse estas consecuencias del hecho de que los debates teóricos sobre la autonomía, al igual que la producción académica sobre el cambio educativo, hayan volcado sus energías en los aspectos formales y procedimentales: cómo se genera el cambio, las capacidades de desarrollo individual e institucional, la estructura organizativa y los procesos de la institución que aprende, etc. (Hargreaves y Hopkins, 1991; Holly y Southworth, 1989). El precio ha sido el abandono del contenido pedagógico, esto es de la cuestión sobre qué prácticas pedagógicas, dirigidas a qué fin (Mitter, 1997). Ello ha conducido a la legitimación de la autonomía del profesorado como una cuestión de procedimientos sin contenidos y sin marco socio-político. En donde por ejemplo, como se viene quejando Elliott (1998), la investigación en la acción se ha desligado del curriculum y la pedagogía. Y en concreto, de una visión del curriculum entendido como un proceso abierto de experimentación pedagógica dirigida por la aspiración de hacer igualmente accesible a todo el alumnado los recursos culturales de la sociedad. Al convertirla en una cuestión formal, la autonomía se ha convertido en una conceptualización que ignora las auténticas autonomías del profesorado, en una intelectualización que ignora el capital intelectual y la experiencia práctica de a quien se dirige el discurso académico; la invención de procedimientos para el desarrollo de la autonomía del profesorado, como forma de mantenimiento del estatus de los académicos, se hace sin tener en cuenta los propios recursos y tradiciones de los docentes para ello. Que las teorías sobre el cambio y la autonomía hayan perdido el contenido pedagógico y las referencias a los compromisos socio-políticos, como algo más que mero procedimiento es probable que tenga que ver (al margen de otros oscuros intereses que en ocasiones puedan estarse instrumentando) con la propia tendencia académica de búsqueda de "la teoría del cambio", de visiones universales, por encima de las diferencias ideológicas. Pero aquí está el error: en la pérdida de comprensión de que las reformas, las innovaciones, los procesos de cambio, son procesos de tensión política entre diversos intereses, grupos sociales, ideologías y estrategias de acción. La idea de la escuela como unidad orgánica de la acción educativa y del cambio puede ser entendida sólo si también se entiende que esto no está al margen de las diferencias sociales, ideológicas y políticas que pueden representarse entre diversos centros y en el interior de los mismos. Y más, en una institución como la escuela que es en sí misma espacio de contradicciones y conflictos sociales. La escuela es un espacio permanente de conflictividad, por lo que es absurdo esperar a saber "cómo se cambia". En ese espacio permanente de conflictividad y contradicción es donde hay que pensar las posibilidades de defensa de ciertas ideas que pueden argumentarse en su valor pedagógico y social y que entrarán en necesaria conflictividad con otras ideas y prácticas.

Pero si se quiere que el debate permanente y abierto sobre los compromisos educativos y sociales trascienda el ámbito académico, y si se aspira a que éste sea un debate fundamentalmente profesional (en el sentido de que se vertebre con las formas prácticas y los pensamientos subyacentes con que los enseñantes realizan su trabajo), entonces es necesario que los docentes puedan tener el reconocimiento y la posibilidad real de tomar parte activa en la elaboración y desarrollo de políticas educativas, así como en la defensa de ideas educativas (en la red semántica que puebla el imaginario sobre la educación y su finalidad) y en el análisis y la denuncia de las condiciones de enseñanza y sus consecuencias. Quizás lo más necesario y urgente, más que inventar procedimientos de "autonomización del otro", sea apoyar la elaboración de voces públicas en las que la discusión pública, política, sobre la educación no esté instrumentalizada por intereses extraños. Pero ello sólo será posible en el contexto de movimientos de enseñantes, entendidos como fuerzas sociales activas, movilizadoras de prácticas e ideas. Y entiendo que aquí el trabajo académico no puede ser otra vez el de tecnologizar otro procedimiento, ahora el de articulación de movimientos. Supongo que no cabe ese riesgo, ni ese error de apreciación. Porque los movimientos sociales se producen en las prácticas y en los compromisos concretos, en la discusión de ideas dentro de colectivos que buscan en su propia práctica profesional dar sentido a su trabajo, expresar públicamente tales motivaciones y las formas en que la enseñanza escolar se encuentra atrapada en conflictos que no siempre son educativos, aunque se proyecten sobre ella.

Notas

Conferencia presentada al coloquio Pedagogia y Autonomia. Universidad del Minho (Braga-Portugal) Braga, 23 de Noviembre de 1998

1. Y produciéndose también una especie de ahistoricismo, en una suerte de —parafraseando a Ball (1997)— post-90ismo, haciendo un vacío de cualquier experiencia y tradición renovadora anterior, como si todo hubiera empezado de nuevo a partir de entonces. Igualmente, se ha producido una dislocación que hace creer que no existe otro lugar fuera de la LOGSE, que no existen experiencias alternativas al margen de lo que supone y propone la reforma oficial.
2. En el caso español todos estos cambios han venido acompañados de una nueva estructura del sistema educativo (cuyo principal aspecto ha sido la ampliación de la escolaridad obligatoria) y de la potente introducción de una concepción constructivista del aprendizaje. Este último factor se ha autojustificado por la pretensión de que la reforma fuera no sólo estructural, sino también metodológica (o como se ha dicho, psicopedagógica), esto es, que afectara a los modos de enseñar. Pero lo que indudablemente ha supuesto también ha sido el establecimiento de toda una estrategia de legitimación de la propia reforma, al ser presentada como una cuestión técnica y científica, y no política, ocultando, en su insistencia en el modo de enseñar, otras decisiones respecto a los contenidos de la enseñanza, o los controles sobre la práctica docente, en forma de una remozada (al

buscar su justificación en el cognitivismo, en vez de en el conductismo) pedagogía por objetivos. Además, ha servido, como producto de los mecanismos de justificación de la reforma que se han puesto en marcha de esta manera, para la introducción y consolidación de un campo disciplinar y profesional: la psicopedagogía.

3. Véase, por ejemplo, (Boyd-Barrett, 1995), en donde se analizan los casos de Inglaterra y España en cuanto que tradiciones curriculares opuestas que, sin embargo, como producto de sus actuales políticas educativas en cada país se han acercado notablemente.
4. Una aproximación diferente a esta cuestión es la que diversos autores están realizando a partir de lo que han llamado "modernización reflexiva". Véase (Beck et al. 1997)
5. Sirva de ejemplo la forma en que un alto ejecutivo español de empresas multinacionales del automóvil, López de Arriortúa (conocido como "Superlópez", por su capacidad para dirigir las empresas en las que trabaja hacia altas cotas de productividad con los nuevos modos de gestión del trabajo), acaba sus conferencias en las que expone sus teorías empresariales: «Para concluir la conferencia López presenta una diapositiva con un pasaje africano y en su mensaje final explica que cada mañana en África, al salir el sol, se despierta una gacela que piensa que tiene que correr más que el león si quiere llegar viva al final del día. El león al despertarse también piensa que tendrá que ser más rápido que la gacela, si no quiere quedarse con el estómago vacío. No importa ser gacela o león, dice López, "lo importante es estar ya corriendo cuando sale el sol"» (El País, 28/5/94). Algo llamativo de este ejemplo, no se sabe si sólo como juego de retórica o si como puro cinismo, es la celebración de la ley de la selva; es la forma de presentar el hecho de comerse unos a otros como una cuestión casi deportiva en la que, como rezaba una vieja campaña española de promoción del deporte, "lo importante no es ganar, sino participar".
6. Esta progresiva tecnologización de la educación no se refiere solamente a los efectos de la aplicación de las nuevas tecnologías y a los intereses que pueda haber tras ello. Tiene que ver también con el desarrollo creciente de programas a aplicar a diferentes órdenes de la práctica educativa: habilidades sociales, educación en valores, programas de tutorías, etc., tecnologización que se mueve en el terreno de la descomposición de aspectos cruciales de la práctica educativa que desaparecen de la relación pedagógica cotidiana y se recuperan como programas específicos, elaborados por profesionales específicos. De esta forma, se disocian de la práctica educativa de los enseñantes aspectos cruciales de su trabajo que ahora deben ser pensados como programas específicos en los que deben ser adiestrados por los nuevos profesionales de la psicopedagogía. Es, paradójicamente, una descualificación que aparece como recualificación.
7. Por supuesto, hay que pensar que en muchas ocasiones todo ello no representa sino la posición ideológica de quienes defienden

tales juegos de lenguaje y tales propuestas (véase como ejemplo, D. Hargreaves, 1994). Sin embargo, en muchas ocasiones, la argumentación a favor de ciertas políticas educativas tiene otros puntos de partida más zafios e intelectualmente más lamentables: desde la inconsciencia de la significación política de las propuestas, pensando que se trata sólo de cuestiones técnicas inocuas, hasta la aceptación de lo que se presenta como una moda, como la forma de pensar y de hacer las cosas ahora, hasta el realismo pragmático de que las cosas son así y más vale que las aceptemos, pasando por el interés gremial o personal, en la medida en que muchas de estas reformas suponen el afianzamiento de un papel social como protagonista de su saber especializado, experto necesario, y que tiene consecuencias en el grado de influencia de las instituciones a las que pertenece o en el beneficio personal (económico y de prestigio).

8. Por supuesto que el análisis de Dubet y Marticcelli, referido a Francia, conserva diferencias muy notables respecto de otros casos. Ellos están hablando de la evolución sufrida por un sistema de escuela republicana y laica, mientras que, evidentemente, en el caso español tendríamos que hablar de la escuela en una dictadura y con una clara ideología nacional-católica. Sin embargo, para lo que aquí nos interesa, creo que la evolución que señalan se corresponde claramente con la seguida por los sistemas educativos en general en los últimos 50 años. Lerena (1980), refiriéndose al caso español (y con otras claves y preocupaciones en su análisis) ha hablado de escuela tradicional-elitista y escuela tecnocrática de masas.
9. Este apartado, y parte del siguiente es un extracto del último capítulo de mi libro La autonomía del profesorado (Madrid: Morata, 1998)
10. Recuérdese que en los años 70 las experiencias de autogestión se referían precisamente a esto, a proyectos de escuelas cooperativas que proponían sistemas de gestión democrática que constituían tanto alternativas de organización como experiencias en sí mismas educativas. Sirva de ejemplo de una experiencia de este tipo (Lara y Bastida, 1982).
11. Este es, en definitiva, el núcleo fundamental de lo que supone el post-fordismo: conseguir la adhesión y la lealtad de los trabajadores haciéndoles participes y creativos en la mejora de los procesos de producción, sin discutir los objetivos de la empresa. Y por supuesto, sin discutir a quién pertenecen los beneficios (la plusvalía), ni la política de personal, ni la flexibilización laboral…

Referencias

Allen, J., Braham, P. y Lewis, P. (ed.) (1992) *Political and Economic Forms of Modernity*. Cambridge: Polity Press (en asociación con Blackwell Publishers y The Open University Press).

Alonso, L. E. (1997) "Universidad pública, no tan pública." El Viejo

Topo, Nº 109, julio-agosto, págs.

Angulo Rasco, J. F. (1992) "Descentralización y evaluación en el Sistema Educativo Español. Algunas claves para el pesimismo." *Escuela Popular*, Nº 3, págs. 23- 87.

Angus, L. (1994) "Sociological Analysis and Education Management: the social context of the self-managing school." *British Journal of Sociology of Education*, Vol. 15, Nº 1, págs. 79-91.

Apple, M. W. (1998) "Are Markets and Standards Democratic?." *Educational Researcher*. Vol. 27, Nº 6, págs. 24-28.

Arnaus, R. (en prensa) "La formación del profesorado: un encuentro con la complejidad educativa." En Pérez, A., Barquín, J. y Angulo, F. (eds) *Desrrollo profesional del docente: política, investigación y práctica*. Madrid: Akal.

Ball, S. J. (1993) "Culture, Cost and Control: Self- Management and Entrepreneurial Schooling in England and Wales." En Smyth, J. (ed.), *A Socially Critical View of the Self-Managing School*. Londres: The Falmer Press, págs. 63-82.

Ball, S. J. (1997) "Policy Sociology and Critical Social Research: a personal review of recent education policy and policy research." *British Educational Research Journal*. Vol. 23, Nº 3, págs. 257-274.

Beck, U., Giddens, A. y Lash, S. (1997) *Modernización reflexiva. Política, tradición y estética en el orden social moderno*. Madrid: Alianza.

Bottery, M. (1992) *The Ethics of Educational Management. Personal, Social and Political Perspectives on School Organization*. Londres: Cassell.

Bowe, R., Gewirtz, S. y Ball, S. J. (1994) "Captured by the Discourse? Issues and concerns in researching 'parental choice'." *British Journal of Sociology of Education*. Vol. 15, Nº 1, págs. 63-75.

Boyd-Barrett, O. (1995) "Educational reform in Spain and in the UK: a comparative perspective." En Boyd- Barrett, O. y O'Malley, P. (ed.), *Education Reform in Democratic Spain*. Londres: Routledge, págs. 65-78.

Brennan, M. (1993) "Reiventing Square Wheels: Planning for Schools to Ignore Realities." En Smyth, J. (ed.), *A Socially Critical View of the Self-Managing School*. Londres: The Falmer Press, págs. 83-99.

Contreras, J. (1996) "Teoría y práctica docente." *Cuadernos de Pedagogía*. Nº 253, págs, 92-100.

Contreras, J. (1997) *La autonomía del profesorado*. Madrid: Morata.

Dubet, F. y Martuccelli, D. (1998) *En la escuela. Sociología del aexperiencia escolar*. Buenos Aires: Losada.

Elliott, J. (1998) The Curriculum Experiment. *Meeting the Challenge of Social Change*. Buckingham: Open University Press.

Fullan, M (1998) "The meaning of Educational Change: A Quarter of a Century of Learning." En Hargreaves, A. et al. (eds.) *International Handbook of Educational Learning*. Dordrecht: Kluwer, págs. 214-228.

Gentili, P. (1997) "El consenso de Washington y la crisis de la educación en América Latina." *Archipiélago*, Nº 27, págs.

Gimeno Sacristán, J. (1994) "La desregulación del curriculum y la autonomía de los centros escolares." *Signos. Teoría y práctica de la educación*, Nº 13, págs. 4-20.

Ginsburg, M. B. (ed.) (1991) *Understanding Educational Reform in Global Context. Economy, Ideology and the State*. Nueva York: Garland.

Hargreaves, A. (1997) "Cultures of Teching and Educational Change." En Biddle, B. J. et al. (eds.) *International Handbook of Teachers and Teaching*. Dordretch: Kluwer, págs. 1297-1319.

Hargreaves, A. y Goodson, I. F. (1996) "Teachers' Professional Lives: Aspirations and Actualities." En Goodson, I. F. y Hargreaves, A. (eds) *Teachers' Professional Lives*. Londres: Falmer Press, págs. 1-27.

Hargreaves, D. H. (1994) "The New Professionalism: The Synthesis of Professional and Institutional Development." *Teaching & Teacher Education*. Vol. 10, Nº 4, págs. 423-438.

Hargreaves, D. H. y Hopkins, D. (1991) *The Empowered School. The management and practice of development planning*. Londres: Cassell.

Hatcher, R. (1994) "Market Relationships and the Management of Teachers." British Journal of Sociology of Education, Vol. 15, Nº 1, págs. 41-61.

Holly, P. y Southworth, G. (1989) *The Developing School*. Londres: The Falmer Press.

Labaree, D. F. (1992) "Power, Knowledge, and the Rationalization of Teaching: A Genealogy of the Movement to Professionalize Teaching." *Harvard Educational Review*. Vol. 62, Nº 2, págs. 123-154.

Lara, F. y Bastida, F. (1982) *Autogestión en la escuela. Una experiencia en Palomeras*. Madrid: Popular.

Lerena, C. (1980) *Escuela, ideología clases sociales en España*. Barcelona: Ariel, 2ª ed.

Lima, L. C. y Afonso, A. J. (1995) "The Promised Land: school autonomy, evaluation and curriculum decision making in Portugal." *Educational Review*, Vol. 47, Nº 2, págs. 165-172.

Lundgren, U. (1989) "Educational Research and the Language of Education." En Ball, S. J. y Larsson, S. (ed.), *The Struggle for Democratic Education. Equality and Participation in Sweden*. Barcombe, Lewes: The Falmer Press, págs. 191-210.

Mitter, W. (1997) "School Autonomy in the Light of Controversial Views: A Comparative Approach with Special Regard to Pedagogical and Curricular Issues in the European Context." En Watson, Modgil, C. y Modgil, S. (eds.) *Power and Responsability in Education* (Vol. 3 de *Educational Dilemmas: Debate and Diversity*). Londres: Cassell, págs. 291-303.

Neave, G. (1998) "The Evaluative State Reconsidered." *European Journal of Education*. Vol. 33, Nº 3, págs. 265-284.

Pérez de Lara, N. (1998) *La capacidad de ser sujeto. Más allá de las técnicas en Educación especial*. Barcelona: Laertes.

Popkewitz, T. S. (1994) *Sociología política de las reformas educativas*. Madrid: Morata/Paideia.

Robertson, S. L. (1993) "The Politics of Devolution, Self-Management and Post-Fordism in Schools." En Smyth, J. (ed.), *A Socially Critical View of the Self- Managing School*. Londres: The Falmer Press, págs. 117-136.

Robertson, S. L. (1996) "Teachers' Work, Restructuring and Postfordism: Constructing the New 'Professionalism'." En Goodson, I. F. y Hargreaves, A. (eds) *Teachers' Professional Lives*. Londres: Falmer Press, págs. 28-55.

Robertson, S. L. (1997) "Restructuring Teachers' Labor: 'Troubling' Post-Fordism." En Biddle, B. J. et al. (eds.) *International Handbook of Teachers and Teaching*. Dordretch: Kluwer, págs. 621-670.

Rudduck, J. (1994) "Reflexiones sobre el problema del cambio en las escuelas." En Angulo, J. F. y Blanco, N. (eds) *Teoría y desarrollo del curriculum*. Archidona, Málaga: Aljibe, págs. 385-393.

Smyth, J. (1993) "Introduction." En Smyth, J. (ed.), *A Socially Critical View of the Self-Managing School*. Londres: The Falmer Press, págs.

1-9.

Weiler, H. N. (1990) "Comparative Perspectives on Educational Decentralization: An Exercise in Contradiction?." *Educational Evaluation and Policy Analysis*, Vol. 12, Nº 4, págs. 433-448.

Whitty, G., Power, S. y Halpin, D. (1998) *Devolution & Choice in Education. The School, the State and the Market.* Buckingham: Open University Press.

Acerca del Autor

José Contreras Domingo

Profesor Titular de la Universitat de Barcelona

Tel.: (34) 934 03 50 45
Fax: (34) 934 03 50 14

Email: Jose.Contreras@doe.d5.ub.es

Dirección

Departament de Didactica i Organitzacio Educativa
Universitat de Barcelona
Passeig de la Vall d´Hebrón, 171
08035 Barcelona (Spain)

Doctor en Ciencias de la Educación por la Universidad de Málaga (Título de la Tesis Doctoral: *La investigación en la acción y el problema de la autonomía del profesorado*).

Profesor de la Universidad de Málaga desde 1983 a 1992.

Profesor de la Universidad de Barcelona, en el Departamento de Didactica y Organización Educativa, desde 1992.

Ha colaborado en diversas investigaciones, entre las que destaca la «Evaluación de la Reforma del Ciclo Superior de EGB en Andalucía», dirigida por Angel Pérez Gómez y José Gimeno Sacristán, financiada por el Plan Andaluz de Investigación (duración: 1986-91), que fue Primer Premio de Investigación Educativa en la convocatoria de Premios Nacionales a la Investigación e Innovación Educativas 1992, realizada por el C.I.D.E. (M.E.C.).

Ha publicado los siguientes libros:

Enseñanza, Curriculum y Profesorado. Introducción Crítica a la Didáctica. Madrid: Ed. Akal. 1990. (2ª ed., 1994).

Models d'investigació a l'aula. (en coautoría con Angel Pérez Gómez y Félix Angulo Rasco). Barcelona: Universitat Oberta de Catalunya, 1996.

La Autonomía Profesional del Profesorado. Madrid: Morata, 1997.

Ha publicado también diversos artículos sobre teoría del curriculum, sobre profesorado y sobre investigación en la acción.

Asiduo colaborador de la revista *Cuadernos de Pedagogía,* ha coordinado el monográfico de la misma sobre «Tendencias educativas en la actualidad» Nº 253, Diciembre, 1996.

Es también miembro de los Consejos de Redacción de las revistas *Investigación en la Escuela* (Universidad de Sevilla), *Temps d'Educació* (Universitat de Barcelona) y de la revista electrónica *HEURESIS: Revista Electrónica de Investigación Curricular y Educativa* (Universidad de Cádiz).

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

Greg Camilli
Rutgers University

John Covaleskie
Northern Michigan University

Andrew Coulson
a_coulson@msn.com

Alan Davis
University of Colorado, Denver

Sherman Dorn
University of South Florida

Mark E. Fetler
California Commission on Teacher Credentialing

Richard Garlikov
hmwkhelp@scott.net

Thomas F. Green
Syracuse University

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Ernest R. House
University of Colorado

Aimee Howley
Ohio University

Craig B. Howley
Appalachia Educational Laboratory

William Hunter
University of Calgary

Richard M. Jaeger
University of North Carolina--Greensboro

Daniel Kallós
Umeå University

Benjamin Levin
University of Manitoba

Thomas Mauhs-Pugh
Green Mountain College

Dewayne Matthews
Western Interstate Commission for Higher Education

William McInerney
Purdue University

Mary McKeown-Moak
MGT of America (Austin, TX)

Les McLean
University of Toronto

Susan Bobbitt Nolen
University of Washington

Anne L. Pemberton
apembert@pen.k12.va.us

Hugh G. Petrie
SUNY Buffalo

Richard C. Richardson
New York University

Anthony G. Rud Jr.
Purdue University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Jay D. Scribner
University of Texas at Austin

Michael Scriven
scriven@aol.com

Robert E. Stake
University of Illinois--UC

Robert Stonehill
U.S. Department of Education

Robert T. Stout
Arizona State University

David D. Williams
Brigham Young University

---

archives | abstracts | editors | board | submit | comment | subscribe | search

```
PCL XL error

    Subsystem:  KERNEL

    Error:      IllegalOperatorSequence

    Operator:   SetClipRectangle

    Position:   13
```

EPAA
• enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 346 times since May 30, 1999

Education Policy Analysis Archives

Volume 7 Number 18 May 30, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ 85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Demonstrated Actions of Instructional Leaders: An Examination of Five California Superintendents

George J. Petersen
University of Missouri-Columbia

Abstract

This exploratory study focuses on the perceived and actual leadership characteristics and actions of five district superintendents in California who focused on the core technology of education - curriculum and instruction. In-depth interviews were conducted with these superintendents, their principals and members of their boards of education. The selection of superintendents for this study were guided by three criteria: peer recognition as instructional leaders, district demographics and aggregated increases in CAP (California Assessment Program) scores in grades 3, 3&6, and 3 6&8 for the academic years of 1986-87 to 1989-90. Interview responses indicated that superintendents in this study perceived four attributes to be essential in their ability to be successful instructional leaders. These attributes are: (1) Possession and articulation of an instructional vision; (2) the creation of an organizational structure that supports their instructional vision and leadership; (3) assessment and evaluation of personnel and instructional programs; and (4) organizational adaptation. By employing responses given by the superintendents in this study and looking closely at what they articulated as their role in promoting curriculum and instruction as well as the larger

organizational structure a preliminary model of perceived superintendent behaviors was constructed.

To confirm perceptions, actions, and behaviors articulated by the district superintendents, triangulation interviews were conducted with school principals and school board members in each of the participating districts. A 52- item questionnaire was also administered to every principal and school board member in these districts. Responses of these personnel confirmed the articulated actions and behaviors of these superintendents in their promotion of the technical core of curriculum and instruction.

Introduction

This research focuses on the perceived instructional leadership characteristics of several highly effective California school superintendents. What makes the research new is not that it comes from a state widely known for its educational innovation, especially that of its chief school officers. The research is new because it focuses on a growing problem now widely shared by chief school officers in this and other states as they struggle with being behind rather than at the leading edge of school reform across the country.

The superintendents at the core of this study were sure that their districts could make a bigger difference in their students' learning than was common across their region and within the state. And despite the remoteness of their central office from the classrooms in which differences must ultimately be made, they were convinced that there must be things that they could do as leaders that would impact on those classrooms curricular, teaching and testing core. If, as the growing body of literature on middle managers suggested, principals could and should be instructional leaders (Dwyer, 1984; Martin & Willower, 1981; Ogawa & Hart, 1985; Peterson, 1984), they wondered why could and should not they?

Their journey to instructional leadership and ultimately effectiveness was neither easy nor unidirectional. Indeed, in even undertaking the journey at all, they had more than their share of obstacles. Chief of these was: A field of educational leadership rive by politics of pragmatism and those of idealism. On one side of this dogfight stood a large majority of respected scholars and practitioners who asserted that educational leadership is primarily a technical matter. For these leaders, the "behavior-thing" had meaning, and leadership revolved around getting others in the organization to accomplish particular tasks. These leaders encouraged potential instructional leaders to pay attention to matters such as personnel administration, school law, school business management and finance, technology and facilities planning. On the other side of this dogfight, stood a smaller but vocal minority of equally respected individuals who asserted that education leadership is primarily a moral matter. For these leaders, the "vision-thing" had meaning and leadership revolved around getting others in the organization to believe in certain things. So these leaders emphasized that the potential instructional leaders should focus on topics such as ethics and values, covenants and commitments, and educational futures instead.

A field of educational leadership in which instructional leadership was of very low priority. Even as top ranked programs of educational administration strived toward major reform in the training of school leaders, the bulk of these reforms rarely focused on issues in instructional leadership. Indeed, one mid-90's study from the influential University Council of Educational Administration (Pohland & Carlson, 1992), ranked instructional leadership seventeenth out of the top 23 subject matter areas offered at the member institutions of UCEA. Even the widely advocated topic of the eighties, instructional supervision, tied for ninth in this survey.

A field of instructional leadership in which the theoretical base is relatively large but the empirical is small. Indeed, even at the time this research began and sometime well after our pool of superintendents had begun their journey as instructional leaders, there were only a handful of studies to which one could turn for guidance about how a superintendent might think, feel, and behave as an instructional leader. While we reserve here the right to summarize later in this paper the findings of two of the best of these studies (Bjork, 1993; Coleman & LaRocque, 1990; Kowalski & Oates, 1993; Murphy & Hallinger, 1986; Peterson, Murphy, & Hallinger, 1986) and compare and contrast them with our own, suffice it to say this handful of studies stands in sharp contrast to the handfuls of studies that have focused on principals as instructional leaders.

A field in which the small pool of empirical research available had not focused on the thinking, feeling, and action of demonstrably effective instructional leaders. The leaders researched were not chosen for their actual success in promoting student learning as that success is typically judged by their public stake-holders, namely, by some kind of test scores or other hard evidence of learning progress. Nor were they chosen for their demonstrated success with those that they were supposed to lead and, in particular, their school boards, their principals and their teachers. So, even if potential instructional leaders took the findings of these few studies on the superintendent as an instructional leader to heart, these leaders had no firm reason to believe that thinking, feeling, and acting as indicated would decidedly impact on the learning of their students or the development of their public and professional staffs.

This study asked demonstrably effective instructional leaders to reflect on the question, "What is your perception of the district superintendent's role in the promotion of curriculum and instruction? The work presented here is based on an examination of the instructional leadership behaviors and activities of five school superintendents in California.

Procedures

Identifying and Selection of Instructionally Focused Superintendents

Employing both quantitative and qualitative analyses drawn from in-depth interviews and school personnel surveys, the collection of data was conducted in three phases. Phase one consisted of inductive and hypothesis-generating interviews with five district superintendents identified and recommended as instructional leaders (Goetz and LeCompte, 1984). The purpose of these interviews was to explore district superintendent's perceptions of functions and responsibilities they perform in the promotion of curriculum and instruction (Seidman, 1991). Phase two consisted of triangulation interviews (based on responses and domains generated from the phase one interviews) with two randomly chosen principals and one school board member in each district. The third phase of the study consisted of administering questionnaires to all principals and school board members in each of these districts who had been active for a minimum of two years during the CAP measurement period and tenure of the district superintendent. Like the phase two interviews, the surveys were used in order to explore the articulated actions and behaviors of district superintendents. Additionally, systematic review of district documentation was also conducted during the third phase.

Selection of Instructionally Focused Superintendents

The ability to locate "instructionally focused" superintendents is not an easy task. No politically savvy district administrator would ever admit that (s)he was not focused on issues of curriculum, instruction and student achievement, but the managerially reality of the position often forces the district superintendent to concentrate on issues other than instruction (Dunigan, 1980; Hannaway & Sproull, 1978; Pitner, 1979). Therefore the selection process of instructionally focused superintendents took a somewhat deductive approach. An initial list of the names of superintendents perceived to be instructionally focused was guided in part by the recommendations of participants in several pilot interviews and conversations (Bogdan & Biklen, 1992;Seidman, 1991;Dwyer, 1984).

These recommendations were obtained from several sources: Faculty members in the Educational Policy, Organizational and Leadership Studies program at the University of California Santa Barbara who were involved in the administrative certification program; pilot interviews with three district superintendents, one assistant superintendent of curriculum and instruction and two elementary school principals located in southern and central California as well as a lecturer in the Confluent Education program at UCSB who had previously served as an elementary school principal and superintendent. This snowball sampling approach (Bogdan & Biklen, 1992) eventually led to a list of eight superintendents.

While recommendations revealed the names of superintendents, the importance of establishing reasonable quantitative measures of instructional effectiveness was the next step. Two sets of data were examined, demographic data on each district and these districts' performance on the California Assessment Program (CAP) achievement test during the tenure of these superintendents.

District Demographics: To ensure that these districts led by these superintendents were similar in type (urban, suburban, rural), size and student populations served, demographic data were collected utilizing the information from the California Basic Educational System (CBEDS) for the school years of 1985-86 and 1989-1990. Information on total student population, minority student population and percentages, as well as the percentages of limited English speaking students (LEP) and percentages of dropouts for each of these districts were complied. Each district was then contacted and asked to provide the percentages of students graduating and going on to institutions of higher education. Examination of these data revealed that they were similar in size, percentage of minority and LEP students, number of student who did not finish school and students who graduated and went on to two and four year institutions.

CAP Achievement Test: Until 1990, the California Assessment Program (CAP) achievement test was administered annually to students in the third, sixth, eighth and twelfth grades CAP assess a range of school achievement including basic skills, critical thinking and problem solving aligned to the California State curricular frameworks (Heck, Larsen, & Marcoulides, 1990). State ranked percentiles for these grades in the general subjects of reading, word recognition, and math from 1985-86 to 1989-90 for these districts were obtained. A review of these data indicated that five of these superintendents were heading districts that had the largest percentile growth in test scores for the areas of reading and mathematics in grades 3, 3&6 and 3,6&8 for the academic years of 1986-87 - 1989-90 (see Table 1). Of course such scores have been criticized as a sole measure of educational effectiveness, still they have been widely used for research in California schools as a common measure of student learning at the state, district, and school level (Hart and Ogawa, 1987; Murphy, Hallinger, Peterson and Lotto, 1987).

Table 1
School District Characteristics

District	Schools	Student Enrollment	District Structure	CAP Percentile Growth (1986-87 to 1989-90) Grades		
				3	**3&6**	**3,6&8**
1	15	9,174	K-12	110	120	138
2	9	6,069	K-12	112	202	174
3	11	5,541	K-12	37	128	126
4	10	9,108	K-12	53	--	175
5	15	9,527	K-12	79	92	150

Instrumentation

A scheduled standardized interview protocol was developed to ascertain the role of the district superintendent in instructional promotion and responsibilities (Goetz & LeCompte, 1984). Questions were primarily open-ended and were based on literature describing superintendent task behaviors and priorities as well as review of instructional models that have been implemented on a district-wide level. Phase Two: Triangulation interview questions based on the information and domains generated by data gathered in the phase one interviews were used with randomly selected principals and school board members in each district. In order to probe the perception of these district personnel, interview questions were generally worded and left open-ended. Phase Three: The fact that responses of principals and school board members in the phase two interviews corroborated and confirmed many of the perceptions and actions articulated by the district superintendents, a fifty-two item questionnaire was constructed and sent to all principals and school board members in each district. Survey items were primarily based on five point Likert scale. There were some binary and forced choice items as well, which primarily examined duties, roles and responsibilities of school principals and school board members.

Data Collection

All superintendent interviews ranged between one and one half to two hours in length. After each interview session, verbatim transcriptions were prepared from an audiotape.

Interviews of principals and school board members were conducted in person and by telephone. These interviews ranged between fifty minutes and one hour and each interview was audiotaped and verbatim transcripts were also made.

A fifty two item questionnaire based on domains and behaviors articulated in the phase one interviews and confirmed in the phase two interviews was administered to every principal and school board member that had been active for a minimum of two years in each of the five school districts. The questionnaire sample consisted of forty-four school principals and thirty- one school board members, sixty-three out of seventy five total respondents, an eighty four percent response rate, completed surveys.

Data Analysis

It is true that informants can and do give inaccurate and misleading data, even though they are doing their best to be helpful (Dobbert, 1982). The reliance on self-reported data by district superintendents could lead to problems concerning the validity of the information received. Because previous research has indicated weak linkages between organizational levels in school districts this study understood that perceptions of actions or behaviors at one level of the organization may not be shared with other levels (Crowson, Hurwitz, Morris, and Porter- Gehris, 1981; Deal and Celotti, 1980; Hannaway and Sproull, 1978).

Answers to interview questions were placed on summary sheets and matrices and then examined to determine if any relationships were apparent. A two-part domain analysis for each interview was conducted (Spradley, 1979) The analysis included analyzing each interview individually across the questions categories. Once individual interviews had been examined and categorized, responses were put on a domain matrices that examined district responses. This matrix was examined in order to determine if themes or consistency were apparent in the perceptions of the respondents regarding their role and participation in curricular and instructional promotion. The open-ended nature of the questions provided an abundance of data on a number of themes.

All analysis of the personnel questionnaire was conducted using SYSTAT (version 5.0). Three types of analysis were used on the completed surveys. First, descriptive statistics were computed for purposes of summarizing the demographic characteristics of the sample and the ratings for each item appearing on the survey (frequencies, means and standard deviations). Second, Cronbach's alpha coefficients (Crocker & Algina, 1986) were calculated in order to ascertain the degree of internal consistency exhibited by the instrument. Examination of the reliability analysis indicated that the instrument exhibited moderated to strong internal consistency. The overall alpha coefficient was equal to .87. Finally, Pearson Product Moment Correlation Coefficients and Kendall-Partial Rank Correlation Coefficients were calculated to test the overall strength and the relationship of four components of the model of superintendent perceived behaviors in district curricular and instructional promotion.

Results

The five superintendents reported that they were involved in all aspects of decision making in their school districts, but all of them concentrated more energy, time and resources to the technical core of curriculum and instruction. First, they articulated a personal vision for the education of children and through different leadership styles, successfully wove that vision into the mission of their districts Second, through the hiring and replacing of personnel, involvement of school board members, shared decision making and the implementation of various instructional strategies they were able to create an organizational structure that supported their vision and role as instructional leader. Finally, they monitored and assessed the programs and personnel using a variety of hard and soft indicators but always with the objective of making the organization more instructionally sound.

Personal Responsibilities

Superintendents in this study gave examples of functions that they did in order to promote instruction within their districts. These functions are referred to as personal responsibilities and can be defined as functions that are neither initiated by nor deferred to other members within the organization. The responsibilities articulated by the participating superintendents were the establishment of an instructional vision, risk taking, being highly visible, modeling and signaling examples of district valued behavior and acting as a district cheerleader.

Vision

Vision has been defined as a set of professional norms that shape organizational activities toward a desired state (Coleman & LaRocque, 1990). Sergiovanni (1990) defines it as beliefs, dreams and direction of the organization and the building of consensus to get there. The term vision in this study is defined as the personal beliefs about the education of children and the expressed organizational goals and/or mission for the school district to accomplish these beliefs.

Superintendent responses strongly indicated that the establishment of a vision or goals was of paramount importance for the district's success in instruction. When asked about their role in the instructional process and specific things that they did to promote instruction their responses were: "The superintendent has to have the vision and sense of what can be" (Superintendent 1, hereafter S1). "I think my role is to establish the vision for this district and to be sure that everybody that works here assimilates and personalizes this vision" (S2). "The vision is real important because it forms a structure or the platform for every decision you make" (S3). "The superintendent has to be more that a catalyst. He must be the keeper and seller of the vision" (S4). "To secure access to a rich curriculum for all students and support networks to help assure that all youngsters are successful is something that we've tried to permeate in terms of our vision for all students" (S5).

Some of the personal visions articulated by these superintendents were: "To ensure that all students acquire the knowledge, skills, and attitudes essential to become productive members of society" (S1). "My commitment to the public is to provide a quality education for all children and to treat people with courtesy and care" (S2). "All students can learn and it is the responsibility of the school to ensure that they are successful" (S4). "I believe it is the responsibility of the school district that every student has access to quality educational programs and access to be successful in meeting the goals of those programs" (S5).

Though the articulation of a vision was essential at the beginning, vision alone is insufficient to promote academic success. The next essential component was the superintendents' ability to successfully integrate the vision throughout the organization. "You have a vision and you transfer that vision into goals. In a school district, whatever it is that you establish as your goals, should then influence the establishment of district outcomes" (S1).

Taking Risks

Another part of the articulation process was taking risks; not always doing the cautious or safe thing. "If you want to improve you have to be willing to take risks when you believed those risks will lead toward better teaching and more effective learning on the part of students" (S5). The superintendents in this study saw themselves as risk-takers, and expressed a personal responsibility to offer instructional programs that they felt were in the best interest for the students and for the goals of the district. Several of the superintendents recounted events when they either eliminated or expanded programs in the district or dismissed popular principals/administrators knowing initially these decisions would risk support and potentially cause a rift in their relationship with members of the school board.

High Visibility

Personal presence was perceived by these superintendents to do three things: demonstrate teacher support, monitor classroom instruction, and to get a first hand account of what was going on at the various school sites. The superintendents in this study indicated that they enjoyed school visitations and felt that their presence on school sites signaled their support of teachers and what they were trying to accomplish. "I show interest in how kids, in how teachers are teaching and kids are learning, by going to the sites and visiting with the teachers and observing classrooms" (S2). Although they enjoyed visiting schools, superintendents saw school visitation as their opportunity to monitor and evaluate each of the school sites. They were particularly interested in assessing technical core operations and expressed that the only way to know what was really "going on" was to spend a good deal of time walking around, looking, asking questions, and being involved. "One of the things that I sees as of significant importance is visibility. Frequent visits, meetings and interaction with staff. Yesterday I visited every elementary summer school classroom. I didn't stay long, but I went and made contact with each one of the teachers. Some places I just stayed fifty seconds, some places I stayed ten to fifteen minutes, depending on the room, but they're used to that. I never tell them when I'm coming to their campuses. I stop in though and say, "I'm here!" They're not allowed to get on the loud speaker and say that the superintendent is here or anything like that. They can't do that. I want to see the real world and everybody's used to that. And so, I'll hit 1,000 classrooms a year" (S1).

Finally, they saw personal visits to schools as a way of managing and reinforcing district goals by talking with principals and teachers about the various program goals and objectives and seeing first hand if district goals were being reached. "Another thing that I like to do and principals and teachers are aware of this. I always encouraged a room environment that is reflective of the instructional program and that includes the display of student work. So, when I visit a classroom, I go in and look at the student work. Now, if I see student work that is really not according to standard, I'll say to the principal, "Have you been in there and looked at that room?" "Go take a look at it!" They know I'll do that. This lets them know that the instructional goals of the district are important" (S3).

Modeling

"Modeling" and "signaling" in these interviews were terms used by the superintendents to mean the same thing. They can be defined as setting personal examples of district valued behavior. "The keeper of the vision has to signal what is important in the company and you signal them in many different ways. You signal through what you write. You signal through what you say. You signal through what you do" (S1). Though modeling/signaling by the superintendents occurred most often in meetings with senior staff, principals, teachers and parents. It also occurred in the classroom. Superintendents indicated that modeling and signaling were articulated through the meeting agendas, in the types of inservice and speakers offered for the staff's professional development, and the allocation of resources given by the district office in the way of staff development. "By supporting financially the district's efforts to do better for kids, I try to model it in everything that I do. We do a lot of training and a lot of staff development. So, we support teachers so they can learn to be more professionally competent and we drive the agendas to a certain extent by the kind of staff development that we provide" (S2).

Cheerleading

Cheerleading was defined as recognizing and presenting programs, schools and individuals that reflect and encompass the vision and mission of the district. As one superintendent said, "Recognizing islands of excellence," within the district. It consisted of the public promotion of innovations, strategies and persons that were working and succeeding in achieving district goals. Cheerleading most often occurred when the superintendent publicly recognized individuals and groups in district meetings, having them conduct presentations in front of parent groups (e.g., PTA) and the school boards as well as honoring them in district newsletters and the local paper. "I'm going out there to recognize high performance to help people celebrate when we have success. Call attention to success. Identify islands of excellence and acknowledge that" (S2). Creation of an Organizational Structure Supporting Instruction

Superintendents in these districts emphasized that the possession and articulation of a vision and personal actions were essential but not sufficient to successfully promote instruction in their districts. The creation of an organizational structure that facilitated and promoted instruction was paramount in institutionalizing their vision. Responses of the superintendents indicated that this was accomplished through two means. First is management of the organization. The rudiments of this strategy as articulated by the superintendents in this study included: Collaboration with the school board, the hiring, transfer and/or replacement of administrative personnel, working and closely supervising school principals, the creation of a hierarchy of district departments, and personal visits to classrooms. The second method was the employment and use of instructional and assessment strategies. These included the use of the California State Curriculum Framework, district- aligned curriculum, district adopted instructional strategies, and intensive staff development.

Management

In the context of these interviews, management represents district organizational policies and personal supervision of members of the organization by the district superintendent in order to facilitate and achieve district goals.

School Board

Common features among these superintendents were the conditions under which they were hired. All five were recruited by the school board with a mandate to improve the instructional program of the district. They felt that this was a significant factor in their ability to promote their ideas and vision with relative ease and in general encountered minimal amounts of conflict with their boards over instructional issues. Though the membership of the school boards has changed during the tenure of each superintendent, the school boards reportedly have supported the efforts of these superintendents to improve the instructional program. To ensure the board's perpetual support, three of the superintendents regularly send board members to conferences, to observe other districts, and include them in staff development inservices focusing on instructional strategies that are being implemented within the district. When asked about getting the school board to share in their vision of instruction and to underwrite them, each superintendent pointed to the fact that they keep their boards involved and appraised of what is happening in the district and the goals they are trying to achieve.

The superintendents in this study expressed that another benefit of their recruitment by their respective school boards was the significant amount of leeway given them to replace personnel in the district. This freedom permitted the superintendents to do two things: (1.) To put key people in important leadership positions (i.e., assistant superintendents and principals) and (2.) to create a hierarchy of district departments.

Hiring, Transfer and/or Replacement of Personnel

The hiring and placement of personnel was articulated as an essential component to the instructional success of their districts. Each superintendent recounted a time when they felt it necessary to replace a member of their senior staff. There were two primary reasons given for these individuals removal. The first was the inertia of the previous administration in the area of instruction and these individual's participation in the inertia. The second and most common reason was the unwillingness of these people to share in and work toward the "new vision" of the incoming superintendent. "I had a person who I felt was a good manager, but just not a good instructional leader and we moved that person into a job that took advantage of his skills" (S5). Only one superintendent said that he replaced a senior staff member because of incompetence. "After I put in a new team, I fired another district administrator because he was totally incompetent. You have to get rid of the 'gate keepers' when you come in to improve a school district" (S1). All of the superintendents articulated that the role of their principals is to be the instructional leaders at their respective sites. A significant part of this responsibility requires the principal to develop detailed site level plans, active leadership, planning, and participation in all staff development, frequent observation of teachers and grounding teacher feedback in district adopted instructional goals.

Superintendents in this study also commented on the fact that it because this was a different paradigm for several of their "old building and grounds" oriented principals, they found it necessary to replace principals in their districts. One superintendent replaced half of his principals in the past six years, four of them in his first year. The reasons were the unwillingness or inability of these principals to share in and work toward the vision of the superintendent. "I had to change a principal because the instructional leadership at that school wasn't what it was supposed to be and wasn't getting to the point where you could see that it was going to get any better. The individual was a nice guy, a great guy, but just not meeting, just wasn't doing it. Couldn't see it. Didn't understand it. Couldn't grasp it" (S3).

Hierarchy of Departments

The importance of personnel being-in-the-right-place was also made evident when these superintendents spoke about establishing a hierarchy of departments within the district. Each of the superintendents maintained that of all the departments in the district, the instructional department was paramount and that other departments existed to support instruction. In only one district was this hierarchy a formalized district policy, the remaining four districts indicated that there was clear "understanding" by the staff members in the district office. In order to facilitate the time necessary to focus on the technical core, superintendents hired and placed highly competent individuals that shared in their vision to head each of the departments. According to the superintendents in this study, the assistant superintendents heading the non-instructional departments, e.g., business and personnel knew of the hierarchy and therefore were given a reasonable amount of autonomy and authority with key check points which permitted easy monitoring by the district superintendents. This alleviated the superintendents from some of the otherwise peripheral organizational concerns and gave them time necessary to promote technical core issues.

Principals

The personal supervision of principals by superintendents was the most common method used to keep a finger on the pulse of district schools. Much of what was said by the superintendents implied that principals were the critical line in the successful promotion of an instructional vision. Principals were required to lead, plan, participate in and act as a resource for teachers at their school site. "We start working on aligning the curriculum and on teaching teachers teaching strategies that would help them to become more effective. We began a very intense program of supervision, evaluation, and feedback for teachers. We taught the principals all this stuff and sent them forth" (S2).

The format of principal - superintendent interaction was fairly standard throughout the five districts. Principals were required to meet with the superintendent on a regular basis. This consisted of between two to four formal meetings a month plus any meetings with the principals at their school site. Each principal was required to write an instructional and leadership plan for his or her school annually. The goals of these plans were to reflect and integrate district policies and objectives with goals for their particular school. These plans were then read and commented on by the superintendent and returned to the principals. In some cases, because of a lack of specificity concerning goals, principals were required to rewrite and resubmit it to the district superintendent.

The school site plans were used in two related -evaluative capacities. The first acted as an assessment tool of the district office in establishing a school's ability to successfully achieve district and site goals outline in the plan. The second was in the evaluation of the principal. All of the superintendent in this study personally evaluated the school principals. By and large, a principal's length of tenure in these districts rested primarily on these evaluations. The evaluations were narrative, detailed and very extensive, "No forms or boxes to check off" (S4). Fundamentally, they were based on the principal's ability to meet the objectives and goals outlined in the school site plan. For example, in one district a goal for each school was to outline and strategically implement the Madeline Hunter Model. The superintendent listened to audiotapes of the principal's conferencing with teachers about the teacher's usage of the model. These conversations then became part of the principal's annual evaluation.

Instructional Strategies

When selecting an instructional model or district wide strategy, there was a consistency across these districts in their criteria. Their decisions were based on three things. First, the model of strategy would have to facilitate the articulated vision and goals of the district. Second, it was necessary that the instructional strategy be grounded in research and practice. Finally, it would have to have a "grass roots" acceptance by a majority of teaching staff. Only two districts made use of the same instructional model, (i.e., Outcome-Based Education and Mastery Learning) while the remaining three used a variety of modes, e.g., Cooperative Learning and Madeline Hunter throughout their schools.

Intensive Staff Development

When a strategy or model had been adopted, extensive staff development was made available to teachers, principals and board members. Each of the superintendents expressed confidence in the professionalism and ability of their teachers but realized that the teachers could benefit from learning alternative ways of presenting material. "I think that we have to let the professionals adapt from a menu of well accepted research and educational practices, and let them use those strategies that best suit them" (S5).

Though each of the districts in this study used a variety of instructional methods, the underlying similarity was that each district made available to their staffs - workshops, conferences, speakers, resources and even courses at local colleges in order to help them to improve their instructional repertoire. One superintendent captured the idea in this statement, "We saw teacher training as an important part of the effort to improve our instructional program. If people know how to teach they will teach. If they don't know how to teach they won't. They'll come up with other things to do to fill the time" (S3).

Assessment and Evaluation

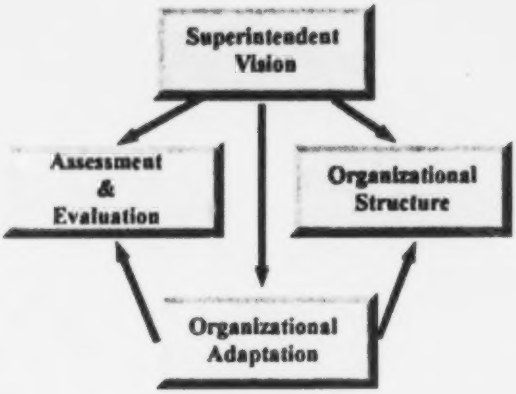
Once a vision had been articulated and programs and personnel were in place, questions such as, "Are the students more successful?" "Is the organization serving the children better?" and "Are programs achieving their objectives?" had to be addressed and answered. According to the superintendents of this study, the next responsibility for the district was to monitor and assess the district's chosen path.

The assessment of instructional success as well as personnel performance relied on the use of both hard and soft indicators. Aside from California Assessment Program (CAP) scores as a means of assessing district and grade level progress in reading, language and math, three of the five districts belonged to the CAS Squared Consortium. CAS Squared made use of an aligned curriculum and provided districts with individual and class scores not measured or reported by CAP. Other evaluative tools included the school site leadership and instructional plans submitted by each school principal at the beginning of the school year. Personal observations by the superintendent and district staff as well as other soft indicators.

A point of interest of this study was the evaluative criteria used by these superintendents in determining whether or not an instructional program should be retained or replaced. The criterion used by the superintendent's was diverse. Three of the districts in this study made use of "soft" indicators when making a decision to retain or replace a program, (i.e., teacher and parent feedback, peer evaluations, community feedback, and district staff feedback) along with some "hard" data, (i.e. CAP scores, district standardized tests, CAS Squared). The two districts using the Outcome-Based Education model made use of "hard" data bands that were tightly aligned to district outcome curriculum goals. If, at the end of one to two academic years, the outcome goals were not being met and or surpassed, the program would be altered or replaced. The underlying criteria in their decisions rested on the idea of whether or not the organization would be able to serve the needs of it's students better. If replacing a program (or person) permitted the organization to improve student learning the replacement generally would be made. "I think, considering everything in the organization, would the total organization be serving kids better or worse? If the bottom line is the organization is going to serve kids better if I make that decision (to replace the program) I'm going to go ahead and do it. If I determine it's not, I'm not" (S2).

Model of Superintendents Perceived Behaviors In District Curricular and Instructional Promotion

By employing responses given by the superintendents in this study and looking closely at what they articulated as their role in promoting curriculum and instruction as well as the larger organizational structure a preliminary model of perceived superintendent behaviors was constructed (See Figure 1).



The model depicts the four significant behaviors these superintendents preformed when promoting instruction within their districts. It demonstrates the flow of their vision and how this vision directs each part of the organizational structure, from the goals and objectives of the district, to the various programs and personnel and the means of evaluation and assessment of both.

Principal and School Board Member's Perceptions

Superintendents stated that principals and school board members played a pivotal role in the successful promotion of instruction within the district. According to the superintendents, principals primarily accomplished this through the writing of school site instructional plans that incorporated district goals and objectives, the observation and evaluation of teachers in the classroom, and planning and participation in staff development and through the monitoring of the principals in these functions by the district superintendent.

School board members (SBM) were encouraged to learn about district instructional strategies in national, state, county and district level workshops and inservices. They were involved in the establishment of district instructional goals and objectives and more significantly the board members that participated in this study articulated an "aligned philosophy" with the district superintendent about what had to be accomplished in order to have an academically successful school district. Other areas of critical importance were fiscal stability of the district and labor peace with certified and classified employees.

Interview and Survey Data

In order to determine whether principals and school board members functioned in the duties and roles as articulated by the district superintendent and what their perceptions of the superintendent are in regard to his role in the promotion of instruction, this study made use of open-ended, triangulation interviews (Spradley, 1979) with ten randomly selected principals and four school board members in these five districts. Confirmation surveys were then designed to corroborated data received from these key informants (Goetz & LeCompte, 1984). The sample of principals and (SBM) surveyed had to have been active in the district for a minimum of two years during the five years of academic growth. The survey sample consisted of forty-four school principals and thirty-one school board members, sixty-three out of seventy five total respondents, an eighty four percent response rate, completed surveys.

Findings

Within district analysis of triangulation interview statements and survey responses with principals and SBM revealed that a significant majority of these pivotal personnel possessed similar perceptions of their role and the role of the district superintendent in promotion of curriculum and instruction. Interviews and within district percentages and frequencies demonstrated that principals perceived themselves as leaders and instructional resources at their respective school sites. (See Table 2)

Table 2
Percent of Principals Answering "Yes" to Survey Questions (n=35)

	Districts				
	1	2	3	4	5
As a Principal were you required to:					
Develop site level leadership plans	100%	88%	100%	100%	100%
Site plans incorporated district objectives	100	100	100	100	100
Regularly observe teachers teaching	100	88	100	100	100
Teacher observations based on district instructional strategies	71	86	100	100	100
Participate in staff development	100	100	100	100	100
Observed by the district superintendent	100	86	100	100	100
Principal evaluations based on goals and objectives developed in site level plan	86	100	75	86	80
Meetings with district superintendent were primarily focused on instructional issues	100	83	100	100	100
Superintendent made frequent school visits	100	100	100	100	100
Superintendent observed teachers teaching	100	100	100	100	100
Superintendent met with teachers at school	100	86	100	100	100
Superintendent is instructionally focused	100	100	100	100	100

Statements and survey responses made it apparent that principals were required by the district superintendent to write site-level plans that incorporated district goals and objectives, to observe and evaluate teachers, to lead and conduct inservices and staff development programs, and to incorporate district adopted instructional strategies in the curricular format at their school sites. Principals were evaluated annually by the district superintendent and a predominant criteria of their summative evaluation was their ability to successfully meet the goals outlined in their school site plans. Principals also articulated and noted that they perceived their respective superintendent as instructionally focused.

School board members (SBM) confirmed much of what was articulated in the superintendent interviews. School board members perceived the district superintendent as instructionally focused and willing to "take risks" in order to promote their instructional vision. They stated and noted a philosophical alignment with the district superintendent on instructional matters, while indicating general involvement in determining instructional goals and objectives for their respective districts. (See Table 3)

Table 3
Percent of School Board Members (SBM) Answering "Yes" to Survey Questions (n=28)

	Districts				
	1	2	3	4	5
As a School Board Member were you:					
Encouraged by the district superintendent to gain knowledge in instructional strategies	63%	100%	83%	100%	100%
Assisted in establishment of district instructional goals	75	100	50	75	100
Overall agreement between SBM and district superintendent in the areas of academic and instructional issues and programs	100	100	83	100	100
Did the district experience labor disputes with staff that interfered with the planning or implementation of classroom instruction?	0	0	0	0	0
Did the district superintendent risk popular support to promote instruction?	88	60	100	86	80
Is the district superintendent instructionally focused?	100	100	83	100	100

They indicated that relationships between the district and certified and classified personnel agencies had not interfered with the planning or implementation of instructional issues during these years of measurement. When queried about the fiscal stability of the district, SBM had stated that the district had become fiscally stable before or under the stewardship of the present superintendent.

As a group, interviews and within district frequencies and percentages indicated that principals and SBM perceived their respective superintendent as possessing and articulating an instructional vision. They also perceived the mission of the school district, the criteria used in the selection and implementation of instructional strategies and staff development as well as the agenda of school board meetings, the criteria used in the assessment of instructional programs as influenced by the vision of the district superintendent. (See Table 4).

Table 4
Percent of Principals and School Board Members(SBM) "Strongly Agreeing" or "Agreeing" to Survey Questions (n=63)

	Districts				
	1	2	3	4	5
Superintendent possessed vision	100%	100%	100%	100%	100%
Vision was focused on instruction	93	92	93	100	100
District mission reflected this vision	100	100	100	100	91
Vision influenced staff development	100	100	93	100	100
Vision influenced instructional programs	86	92	93	100	100
Vision influenced school board agenda	86	100	100	100	91
Vision influenced principal evaluations	100	100	93	100	82
Vision influenced criteria used in assessment of instructional programs	86	92	100	100	100
Vision influenced the modification of district instructional programs	93	100	100	100	100
Superintendent encouraged collaboration	77	92	92	40	100
Superintendent received input from principals	53	100	86	55	70
Superintendent received input from SBM	64	92	65	64	64
Academic success due in part to superintendent vision and involvement	93	92	93	100	100
Superintendent strongly focused on curriculum and instruction	100	100	100	100	100

Though a majority agreed that the superintendent encouraged collaborative decision making, responses from all districts in this study indicated that collaboration primarily occurred at the school site level with little input from groups such as teachers, principals, and parents at the district level. Principals and SBM perceived that the assessment of instructional programs and their modification relied on both "hard" and "soft" indicators, while the replacement of district and school site personnel relied more on ‘hard’ indices (e.g., test scores, ability to achieve stated goals and objectives.) Participants also indicated that the academic success of their respective district could be, in part, to the vision of the district superintendent in instructional matters.

Conclusion

The findings and conclusions of this study are limited in their generalizability since they were derived from exploratory interviews and survey instruments and were only used in five non-randomly selected medium sized school districts in California. The explanation and interpretation of the findings also has several reasonable alternative explanations. While superintendents in this study credit personal vision as fundamental to the instructional success of the district, there are at least three important organizational factors that may serve as reasonable alternative explanations for these districts success. They are: 1.) The ability of the superintendents to replace principals and other administrators who did not share the superintendent's vision and mission. 2.) The fiscally stable conditions of the district as well as the latitude given each of these superintendents by their boards of education. 3.) The strict alignment of the district curriculum to teaching strategies and district outcomes. Limitations also reveal that further general research is recommended in order to obtain a more complete comprehension of the superintendent's role in curriculum and instruction.

With this caveat aside, the findings from this study suggest a new and somewhat different leadership role for the district superintendent in the core technologies of curriculum and instruction. Emerging from the data were several critical themes demonstrating consistencies among the instructionally focused superintendents. This included creation of a vision, increased visibility, modeling of academic expectations, developing rapport with the school board, and management of instructionally oriented programs.

First, this study demonstrates the importance of creating an instructionally oriented vision and communicating this vision throughout the school district. For example, each of the superintendents in this study demonstrated an instructionally oriented vision for academic success. This finding is consistent with other research that suggests that educational reform is impossible without visionary leadership by superintendents (Kowalski & Oates, 1993). These superintendents communicated their vision of excellent teaching and learning through continual communication with principals. Carter et. al., (1993) describe the importance of utilizing principals to carry their message to each individual school in the district. Superintendents attempted to transform their vision into an instructionally oriented vision for academic success through strong and tightly coupled leadership. Vision and strong leadership has previously been determined to be a critical element of successful instructional leadership (Bredeson, 1996; Carter et al., 1993; Murphy & Hallinger, 1986; Peterson, Murphy & Hallinger 1987).

Second, high visibility was also demonstrated by the superintendents in this study. High visibility in schools and in classrooms has been linked to instructionally effective schools (Bjork, 1993). This visibility also led to the modeling of high academic expectations, which was found to be a critical action demonstrated by the instructionally successful superintendents. This is also consistent with past research that deems frequent visits to schools as a necessary component of demonstrating the importance of instruction (Carter et al, 1993). These superintendents visited classrooms frequently throughout the district and reported classroom observations to the principal. Consequently, the superintendents modeled the importance of instruction to the teachers, students and principals. Perceived discrepancies, by the superintendent, between the districts mission and the teaching in the classroom were quickly disseminated to the principal who could act to correct the differences with the individual teacher.

Third, each superintendent was able to illustrate the importance of instructional leadership through professional development and shared decision-making. Each district made available an abundance of workshops and possibilities of attending conferences promoting alternative teaching methods. This availability of professional development opportunities demonstrated the importance of teaching and learning in the district. Through these visible opportunities for teachers, each superintendent illustrated that teaching and learning was clearly the most important objective of the school district. Through providing such professional development activities the superintendent is communicating the importance of teaching and learning.

The study demonstrated the critical nature of the superintendent's individual action of creating an academic oriented vision and maintaining this vision through high visibility. With each visit to a school the superintendent modeled the importance of the instructional oriented vision through appearance as well as signaling to the principal when discrepancies arose between the district wide mission and an individual teacher's actions in the classroom.

Fourth, each participating school district demonstrated support from the school board for superintendent decision making. In this study, these instructionally focused superintendents had clear support from the school board. In fact, most of these five superintendents were hired due to their previous instructional experience and success. This study supports previous research, which has demonstrated the importance of school board support (Griffin & Chance, 1994). Support of the school boards permitted the superintendents in this study to take significant risks in their promotion of the technical core. This finding has reflects previous research in this (Kowalski & Oates, 1993). Without the support of the school board, a superintendent is less likely to take risks that could yield academic results due to the fear of losing his/her job. With the average tenure of a superintendent currently 2 to 3 years, this is a realistic fear.

Furthermore, school board support is directly related to additional findings in this study. These superintendents were able to exercise power in regard to placement of individuals in positions of leadership (i.e., district administrators and principals) due to the support and freedom in decision making extended from the school board. By allowing the superintendents to place individuals in strategic positions they are guaranteed to align self-chosen individuals to positions that greatly influence instructional leadership. This authority vested by these school boards into their respective superintendents permitted them to replace administrative team members who were not instructionally oriented and/or committed to the instructional vision of the district superintendent.

Previous studies have demonstrated the importance of shared decision making with the superintendency and the school board, yet this study exceeds this interaction with decision making freedom extended to the superintendent. This finding should lead to new research into the dynamics of decision-making freedom for the superintendent and effective schools.

Fifth, each of the superintendents in this study used assessment and evaluation techniques to determine if the district's school performance was meeting articulated expectations. Their employment of curricular designed principal evaluation, feedback from district personnel, standardized test scores and district instructional programs. This information provided the superintendents in this study with feedback mechanisms on the success of their programs. This type of evaluation is consistent with research in this area (Coleman and LaRocque, 1990; Murphy and Hallinger, 1986).

References

Bjork, L.G. (1993) Effective schools-effective superintendents: The emerging instructional leadership role. *Journal of School Leadership. 3*, 246-259.

Bogdan, R.C. and Biklen, S. K. (1992). *Qualitative research in education.* (2nd ed.) Boston: Allyn and Bacon.

Burns, J.M. (1978). *Leadership*. New York: Harper and Row.

Bredeson, P. V. (1996). Superintendent's roles in curriculum development and instructional leadership: Instructional visionaries, collaborators, supporters, and delegators. *Journal of School Leadership, 6*. 243-264.

Bredeson, P. V. and Johansson, O. (1997) Leadership for learning: A study of the instructional leadership roles of superintendents in Sweden and Wisconsin. Paper presented at the annual conference of the American Educational Research Association, Chicago, Illinois.

Campbell, R.F., Cunningham, L.L., McPhee, R.F., and Nystrand, R.O. (1970). *The Organization and control of American schools*. (2nd ed.). Columbus, OH; Merrill Publishing Co.

Carter, D.S.G., Glass, T.E., and Hord, S.M. (1993). Selecting, preparing and developing the school superintendent. (ERIC Document Reproduction Service No, ED 393215).

Coleman, P. and LaRocque, L. (1990). *Struggling to be 'good enough': Administrative practices and school district ethos*. London: The Falmer Press.

Crane, D. D. (1989) The leadership of an insider superintendent with a cosmopolitan orientation in an expanding and changing orientation. Unpublished doctoral dissertation, University of California, Santa Barbara.

Crocker, L. M. and Algina, J., (1986). *Introduction to classical and* *modern test theory*. New York: Holt, Rinehart and Winston.

Crowson, R.L., Hurwitz, E. Jr., Morris, V.C., and Porter-Gehrie, C., (1981). The urban principal: Discretionary decision making in a large educational organization. Chicago. University of Illinois at Chicago. College of Education.

Cuban, L. (1984, May). Transforming the frog into a prince: Effective schools research, policy, and practice at the district level. *Harvard Educational Review, 54* (2), 129-150.

Deal, T.E. and Celotti, L.D. (1980). How much influence do (and can) administrators have on classrooms? *Phi Delta Kappan*. 61, 471-473.

Dobbert, M. L. (1982). *Ethnographic research: Theory and application for modern school and societies.* New York: Praeger.

Duignan, P. (1980, July). Administrative behavior of school superintendents: A descriptive study. *Journal of Educational Administration, 28.* (1), 5-26.

Dwyer, D.C. (1984, February). The search for instructional leadership: Routines and subtleties in the principal's role. *Educational Leadership, 41.* (5), 32- 37.

Edmonds, R. (1979). A discussion of the literature and issues related to effective schools (ERIC Document Reproduction Service No. ED 170394).

Goetz, J.P., and LeCompte, M.D. (1984). *Ethnography and qualitative design in educational research.* Orlando: Academic Press.

Griffin, G., and Chance, E.W. (1984). Superintendent behaviors and activities linked to school effectiveness: Perceptions of principals and superintendents. *Journal of School Leadership, 4.* 69-86.

Hallinger, P., and Murphy, J. (1986). The superintendent's role in promoting instructional leadership. *Administrator's Notebook.* University of Chicago, 30 (6).

Hannaway, J., and Sproull, L. S. (1978). Who's running the show? Coordination and control in educational organizations. *Administrator's Notebook.* University of Chicago. 27 (9).

Hart, A.W., and Ogawa, R.T. (1987, Winter). The influence of superintendents on the academic achievement of school districts. *Journal of Educational Administration, 25* (1), 72-84.

Herman, .J.L. (1990). Instructional leadership skills and competencies of public school superintendents: Implications for preparation programs in a climate of shared governance. Paper presented at the annual meeting of the Southern Regional Council on Educational Administration. (ERIC Document Reproduction Service No. ED 328980).

Hitt, W. D. (1990). *Ethics and leadership: Putting theory into practice.* Columbus: Battle Press.

Jackson, B.L. (1995). *Balancing act: The political role of the urban school superintendent.* Washington, D.C.: Joint Center for Political and Economic Studies.

Kowalski, J., and Oates, A. (1993). The evolving role of superintendents in school-based management. *Journal of school leadership, 3*, 380-390).

Leithwood, K. (1995). *Effective school district leadership: Transforming politics into education.* Albany, NY.: State University of New York Press.

Martin, W. J., and Willower, D. J. (1981, Winter). Characteristics of instructionally effective school districts. *Outcomes, 8* (4), 26-33.

Murphy, J., Hallinger, P., Peterson, K.D., and Lotto, L. S. (1987). The administrative control of principals in effective schools districts. *The Journal of Educational Administration, 15* 161-192.

Myers, M.D. (1992, Winter). Effective schools and the superintendency: Perception and practice. *Contemporary Education, 63*. (2), 96-101.

Ogawa, R.T., and Hart, A.W. (1985). The effect of principals on the instructional performance of schools. *Journal of Educational Administration, 23* (1), 59-72.

Petersen, G. J. (1993). Keeper and seller of the vision: The district superintendent's perceived role in the promotion of curriculum and instruction. Dissertation Abstracts International, 54 (09), 3285.

Peterson, K.D. (1984, December). Mechanisms of administrative control over managers in educational organizations. *Administrative Science Quarterly, 29* (4), 573-597.

Peterson, K.D., Murphy, J. and Hallinger, P. (1987, February). Superintendents' perceptions of the control and coordination of the technical core in effective school districts. *Educational Administration Quarterly, 23* (1). 79-95.

Pitner, N. J. (1979). So go the days of their lives: A descriptive study of the superintendency: A descriptive study of the superintendency. University of Oregon, Oregon School Study Council. ERIC Document Reproduction Service No. ED 165 299.)

Pitner, N. J., and Ogawa. R.T. (1981, Spring) Organizational leadership: The case of the school superintendent. *Educational Administration Quarterly, 17* (2). 45-65.

Pohland, P. A., and Carlson, L. T., (1992, December). Upgrading the cohort: A model for admissions to preparatory programs in educational administration. *Journal of Personnel Evaluation in Education, 6* (2) 123-39.

Salley,C.(1979). Superintendents job priorities. *Administrator's Notebook.* University of Chicago. 28 (1).

Schlechty, P. (1990). *Schools for the 21st century.* San Francisco: Jossey-Bass.

Seidman, I. E. (1991). *Interviewing as qualitative research: A guide for researchers in education and the social sciences.* New York: Teachers College Press.

Sergiovanni, T.J. (1990). *Value-added leadership: How to get extraordinary performance in schools.* San Diego: Harcourt Brace Jovanovich.

Spradley, J.P. (1979). *The ethnographic interview.* New York: Holt, Rinehart and Winston.

Thompson, J. D. (1967). *Organizations in action.* New York: McGraw Hill.

Willower, D. J., and Fraser, H.W. (1979). School superintendents on their work. *Administrators Notebook.* University of Chicago, 28 (5).

Wimpelberg, R.K. (1988, May). Instructional leadership and ignorance: Guidelines for the new studies of district administrators. *Education and Urban Society, 20* (3), 302-310.

Wirt, F. M. (1990) The missing link in instructional leadership: The superintendent, conflict and maintenance, (ERIC Document Reproduction Service No. ED 327945).

Note

An earlier version of this paper was presented at the Annual Meeting of the American Educational Research Association San Diego, California, April 13-17, 1998. I would like to thank James Block, University of California, Santa Barbara and Joseph Slowinski, M.Ed., doctoral candidate at Indiana University, Bloomington, IN for their assistance in the preparation of earlier drafts of this manuscript.

About the Author

George J. Petersen
Email: gjpetersen@iname.com

George J. Petersen, Ph.D. is an Associate Professor in the Department of Educational Leadership and Policy Analysis at the University of Missouri-Columbia. Correspondence concerning this article should be addressed to: George J. Petersen, Ph.D., Department of Educational Leadership and Policy Analysis, 211 Hill Hall, University of Missouri-Columbia, Columbia, MO 65211.

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa- DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Maria Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 230 times since June 4, 1999

Education Policy Analysis Archives

Volume 7 Number 19 June 4, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ
85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

A Statist Political Economy and High Demand for Education in South Korea

Ki Su Kim
Memorial University of Newfoundland
Canada

Abstract

In the 1998 academic year, 84 percent of South Korea's high school "leavers" entered a university or college while almost all children went up to high schools. This is to say, South Korea is now moving into a new age of universal higher education. Even so, competition for university entrance remains intense. What is here interesting is South Koreans' unusually high demand for education. In this article, I criticize the existing cultural and socio-economic interpretations of the phenomenon. Instead, I explore a new interpretation by critically referring to the recent political economy debate on South Korea's state-society/market relationship. In my interpretation, the unusually high demand for education is largely due to the powerful South Korean state's losing flexibility in the management of its "developmental" policies. For this, I blame the traditional "personalist ethic" which still prevails as the *modus operandi* of the agents in the education market as well as state bureaucrats.

The Situation in Question

Remarkable in South Korea's contemporary education is the accelerating speed of its growth. Primary education (Grades 1-6) saw full participation in the late 1950s, about 70 years after its inception in 1886 (Yun et al. 1996, p. 67; MCE 1988, pp. 151-3). From 1960 on, lower secondary education in "middle schools" (Grades 7-9) expanded and reached full participation within 30 years. Then, from 1970 on, upper secondary education in "high schools" (Grades 10-12) followed suit, taking only 20 years to reach a similar state of full participation. (Note 1) From 1990, finally, higher education picked up speed in growth and by 1998 had come to admit 84 percent of the year's high school leavers (see Table 1).(Note 2) Now, thus, South Korea is entering into a new age in which almost all children go through the schooling process as far as universities and colleges.

Table 1
Graduate Advance to Next-Level Schools (%) (MOE 1998, p. 3)

Advance Type	*1970*	*1980*	*1990*	*1998*
Primary to Middle	66.1	95.8	99.8	99.9
Middle to High	70.1	84.5	95.7	99.4
High to University	31.9	44.0	46.0	83.7

What is more remarkable is that the growth of education has been attained largely at the direct financial expense of the students' parents. In 1997, for instance, state grants for private universities and junior colleges covered barely 3 percent of their operation costs although such institutions accommodated 83 percent of all students in higher education. In the same year, state universities too collected over 40 percent of their expenses from tuition and other fees. Even schools for compulsory education (all primary schools and some rural middle schools) raised more than 10 percent of their expenses from their pupils although they did not collect tuition fees as such.(Note 3) The South Korean state, according to Cho (1994, p. 101), "has not given due regard to the importance of education while concentrating its energy on economic growth." The impressive growth of education, in his view, is "the result of the natural course of events" in the education market, in which high demand constantly called for increased supply (Ibid., 101n). South Koreans' demand for education is indeed high and firm is their determination to get education for their children. Evidence is the intensity of entrance competition which drives students in all stages of schooling (even preschool) to costly neighbourhood cramming classes.(Note 4) According to South Korean studies (Kong et al., 1994; Kim YC et al. 1997), parents' miscellaneous educational expenses, of which cramming fees occupy the largest portion, far exceed the costs for operating the nation's entire education system. Such was the situation in 1998 that Kim Dae Jung, the incumbent president installed in February of that year, vowed, as his predecessors did, to solve the entrance competition question as his top priority policy task.

Why is there in South Korea such a high demand for education?

Existing Answers

Possible answers may well be, and indeed the literature points out, that there is in South Korea a culture of valuing education, that the recent economic development has greatly enhanced South Koreans' purchasing power, and that a successful life in their society requires a high level of schooling. While sensible generally, these oft-given answers possess limited persuasiveness.

The point of departure of the cultural account is the phenomenon of "excess demand" in education commonly found in East Asian nations. Evidence is high application rates for entry, low success rates and a large number of repeaters, as a study of the Japanese case aptly puts (James and Benjamin 1988, 56). Since such phenomena are uncommon in non-East Asian societies, it would be fair to associate them to what is unique to East Asia, the culture—the "Confucian continuum" of social life—which respects educated persons (e.g., Tu 1997; McAdams 1993; Smith 1991; Song 1990; Whitehill 1987). Since education earns respect in this culture, anyone desirous of a respectful life would surely seek it. Non-East Asian writers here often jump on to relating East Asia's recent economic rise to that culture, which has transformed the masses to an educated workforce. East Asian writers seldom agree to this favourable observation. While admitting the benefits of the culture, they are generally cynical as to its effect, obviously because they are so familiar with the problems entailed by the culture's thrust for education.They call this thrust "*kyoyungnyol*" (or "*kyoikunetsu*" if they are Japanese), which means a blind enthusiasm *cum* fever for more and higher education. In their view, this cultural thrust generates more evil than good (e.g., Kim YH, et al. 1994; Park NK 1994, and Kim IH 1991).

Let the Confucian culture be more prone to valuing education than other cultures, and let it be more so in South Korea than elsewhere in East Asia. Neither of the cultural accounts actually takes us beyond re-identifying what has already been identified, that is, that there is a high demand for education in a society where Confucian culture remains entrenched.(Note 5) The socio- economic accounts, meanwhile, sound better because they relate the phenomenon to what has enabled the consumers to demand more and higher education, and for what practical purposes. Yet a few points need be sorted out. The enhanced purchasing power account,(Note 6) to begin with, is acceptable in the sense that demand for education would not rise unless money was available for purchasing it, given that education in South Korea was not free. True, South Korea's economic development has greatly improved household finance. Yet economic development may not explain the phenomenon of 1998 unless its fruits had been distributed as equally as to enable almost all South Koreans to pay for education.(Note 7) This is not the case, however. Unequal distribution of the fruits has already caused several social turmoils during the last few decades. Given this, it may be rather reasonable to suspect that the price of higher education, tuition fees in particular, has been kept low enough for most parents to believe that they could somehow afford it if they made self-sacrificial commitment, or that somehow they have been compelled to pursue more and higher education for their children, or both. In a country where demand in education is high and educational institutions seldom have revenue sources other than tuition fees, it is unlikely that those institutions have voluntarily kept their fees low. They might do so, however, if coerced by a certain extra-market element, the state in particular.

The account concerning the purposes of seeking education addresses the society which South Korean scholars characterize by the word *hangnyogchuyi* (just as the Japanese typify theirs by *gakurekishugi*, the same word pronounced differently). This refers to a method of assessing individuals' merit by considering the last school they have attended, preferring university graduates to high school graduates and graduates from a prestigious university to those from other universities. This method is so widely employed that not only the employer who interviews job applicants but also the mother who looks for a bridegroom for her daughter, rely on it. Since *hangnyogchuyi* is widespread and a university degree is "a ticket to the social elite" (Pak 1997), the argument goes, young people cannot but pursue higher education preferably at a prestigious university so long as they desire entrance to a high and respectable echelon of society (Sorenson 1997; Lee and Brinton 1996, 177-92; Ryoo 1995). Even if it is granted that the practice of *hangnyogchuyi* is more salient in South Korea than elsewhere, none the less, that does not explain the high demand for education, because the practice can as well be an effect of the high demand for education and its competitive pursuit, as it can be the latter's cause. When the great majority of a society's membership competitively seek higher education at a prestigious university, then *hangnyogchuyi* must actually be a dependable way to selecting employees or sons-in-law.

Although the cultural and socio-economic accounts are useful, their weaknesses may not be overlooked. A prominent weakness is their focus on the demand side of the education market and leaving the supply side out, as if the demand side is self-sustaining. Yet in the education market, as in any market, demand is usually interdependent with supply. Its rise becomes sooner or later balanced out by the rise in supply, and its fall results in the fall in supply. So long as this rule stays, demand in education may not persist to rise if supply is left to fluctuate freely. Whereas, if supply is deliberately kept low, demand can surge unusually depending on the nature of the commodity at stake. The situation of South Korea's education market in 1998 was that university entrance competition was intense (meaning that demand was very high) even though universities and colleges had prepared sufficient student places to admit some 84 percent of the year's high school leavers. As a matter of fact, supply in that year was raised well in excess of demand because post-secondary institutions as a whole ended up filling only 94 percent of their capacity (MOE 1998, 3). Even so, entrance competition was as intense as before. If South

Korea's education market was not an exception to the general market rule, it would be sensible to assume that, here again, a certain extra-market element was deliberately creating "poverty in abundance" on the supply side. It could, for instance, be making entrance to certain universities more difficult than others, and it could be keeping the price of higher education relatively low to motivate massive applications.

Here, the unusual rise of demand in South Korea's education market seems to warrant an assumption that it is not a "result of the natural course of events" in the education market but rather a market phenomenon due to intervention by certain external forces, especially the state—a state which is in fact well noted for heavy market intervention. In the rest of this article, I shall explore a new interpretation of the phenomenon by pursuing this assumption. In doing so, I shall critically borrow from some of the theses that have emerged from the recent political economy debate on the relationship between South Korea's state and society/market.

The Political Economy of the Developmental State

(1) The "Personalist Ethic" and the Statist Account

The political economy debate on South Korea's state-society/market relationship has been triggered by the fact that the country's society was not as "civil," nor its state as "political,"(Note 8) as in Western nations, while its economy achieved incredible development under the state's market intervention. Conspicuous in the not-so-civil society is the prevailing pre-capitalist *modus operandi*, in which business transactions are made not by free and open competition as in more "capitalist" societies, but by such personal connections as clanship, alumni fraternity, regional ties and bribery (which is a method of creating new ties and consolidating existing ones). Free and open market competition is also severely restricted by the state which controls transactions and allocates important resources. What further complicates the matter is that the intervening state has been under the control of the military elites who, having usurped power by coups d'état, did not abide by public demand and often acted against it. Their governments were less likely than elected governments to function "politically" in the interest of the public. Even so, surprisingly, South Korea's economy has grown to be the eleventh largest in the world. How could this happen?

Some writers seek answers in what appears to be an obstacle to economic development, that is, the traditional *modus operandi* and the state's market intervention. The reason for considering the traditional *modus operandi* to be obstructive is that it leads supply to meet demand by collusion. This type of transaction may well be suitable for in-group distribution of available resources, but not for promoting the wide production of resources and their consumption, while the latter are essential for economic development. Unlike free and open market competition, collusion precludes incentives for such activities. Dore (1980 and 1982) and Chang (1991) side step this commonplace observation by viewing the traditional modus operandi—"the personalist ethic" as they call it—in terms of the small *groups* which the colluding persons form, such as the bodies of clan, alumni, and so on, rather than in terms of individual collusions as such. Thus viewed, the semi-capitalist South Korean market appears to be operated by various personally-tied groups, each of which seeking its own group interest collectively internally and competitively externally. Although these groups differ from the "individuals" in other markets, the argument goes; they are legitimate parties in the South Korean market and are as active as individuals elsewhere. Therefore, they conclude, the personalist ethic is as valid a "capitalist" *modus operandi* as the individualist ethic.(Note 9) And if the market which is pervaded by the personalist ethic has grown to be a world-class economy, they argue, this ethic appears to have worked as efficiently in South Korea as the individualist ethic elsewhere.

Market intervention by the state—a "statist state" as it may be called because its interests override private interests or a "developmental state" because it controls the market in order to make it "develop"—has drawn attention from most writers on South Korea's political economy although they are divided over whether to recognize the state's contribution to economic development. Of focal concern here is the period between 1961 and 1979 during which General Park Chung Hi's military regime vigorously pursued economic development. For this purpose, the developmental state planned the economy's development, rigorously invested in the public sector, and supplied private business with massive financial support to promote industries (Lee YH 1997, p. 1). In short, it managed the economy with the specific purpose of promoting planned development. Writers of the "statist"viewpoint—statist, this time, because they emphasize the state's leading role in economic development—link this type of state intervention directly to the economic development that has actually taken place. They thus claim that, had it not been for the developmental state's effective distribution of resources for their maximal utilization, the economy would not have developed as it did (e.g., Amsden 1994).(Note 10)

The personalist and statist theses, thus, emerge as potential explanations for the educational development of South Korea, because should they explain economic development there would be no reason for their not doing the same to the development of education, which also bears economic features. The theses, however, may not squeak by without reservation. First, the state which intervened in the market with the political agenda of economic development was not free from the personalist ethic which prevailed in the market. Government leaders and state bureaucrats very often operated according to personal ties, such as alumni connections, regional loyalties, military collegiality and bribery. Even state policies sometimes noticeably favoured certain regions and groups while disfavouring others. Important resources, such as loans from the state-controlled banks, foreign loan guarantees and licences for monopoly businesses, were very often allocated through such personal connections. The *political* state of South Korea, therefore, functioned very often as a mere *economic* agent, seeking private interest while intervening in the market in the name of public interest, thus obfuscating the state-society distinction.(Note 11)

This leads to another reservation. If the state in its operation was affected by the personalist ethic while politically committed to promoting economic development, to that extent, one may say, its political commitment to economic development was impaired by the private, economic interests of the personalist market. In this situation, the state's political drive for economic development can hold sway only to the degree that it is unaffected by the personalist ethic. Beyond this degree, the developmental policy which the state pursued might in fact be a personalist manipulation of the economy. Given that the personalist ethic was prevalent in society and even state officials and government leaders were affected thereby, the degree could not be great. Put boldly, the state's political intervention can be rather more exceptional than normal. If so, saying that South Korea's economic development was due to the developmental state's intervention, may be tantamount to saying that the economic development was due to the personalist ethic which held sway over both society/market and the state. The theses about the personalist ethic and the state's leading role in economic development are not separate, therefore.

Can, then, the personalist ethic take credit for South Korea's economic development? The affirmative suggestion by Dore and Chang would be acceptable if the groups emerging from personal ties indeed *competed* with each other for important resources. Yet if they did so, the personalist ethic would no longer be the groups' *modus operandi* because, in spite of their grounding in personal ties, they were actually operating in the *modus operandi* of the individualist ethic, that is, *competition*, as in Dahl's (1967) pluralist society. The groups tied by personal connections did not, however, function *as groups* in the pursuit of important resources, nor did they *compete* for such resources. Although personal ties often developed to alumni associations, societies of regional origin, and the like, it was not the organizations themselves but some members of the organizations who actually reaped benefits. They did so because the state's allocation of important resources was frequently made via *personal* collusion and seldom via competition among organizations. This practice had to do with the lack of popular support for the military elite, not only for the initial Park regime which was born out of the 1961 coup but for the subsequent Chun Du Hwan and Roh Tae Woo regimes which emerged out of another coup in 1979 and lasted until 1992. Through the years of their domination, the military elites sought to ground their support in personal ties rather than in the popular vote (which was simply not there) and allocated crucial resources to that end. This kind of practice was common in the private sector as well, where chaebols, whose monopoly in the domestic market the state endorsed, often awarded contracts to minor enterprises in the same way. The personalist ethic may not claim credit until this mode of transaction is proven to be more effective than free and open competition in the utilization of important resources for economic development.

(2) "Structural Problems" and the Institutionalist Account

In fact, there are grounds for more persuasively arguing that the personalist ethic of South Korea's pre-capitalist market has impeded economic development as well as the state's own developmental effort. An exemplifying case for this is the economic crisis which began in November 1997.

The crisis broke out when foreign capital pulled out of South Korea, which immediately placed the debt-ridden chaebols on the verge of bankruptcy and activated a disastrous chain reaction in the economy under their domination. While controlling major industries, the chaebols relied heavily for their operation on the state's favourable financial support and protection of monopoly rights in the domestic market. On the safe grounds provided by state support and protection, the chaebols quickly grew, yet at the same time they lost the ability to survive market competition. Their business organizations invariably grew to be big, bureaucratic and inefficient. In spite of protection, therefore, most of them repeatedly incurred losses; and such losses were made up for by additional favourable loans from state-controlled banks. Justification of such additional loans required further expansion in business operations. The more chaebol businesses grew in size, thus, the more dependent they became on such favourable loans. Then, the sudden depletion of financial resources brought this practice to a halt because it took away from the "strong state" (Kim EM, 1997) the very resources for favourable allocation. The crisis glaringly demonstrates that South Korea's economic development was not development in a solid market economy but growth in the volume of the economy due to the expansion of chaebol monopoly businesses under state protection and support. Yet while the economy grew, "structural problems" continued to be exacerbated, such as: the stifling of private initiatives under state control and chaebol monopoly, the inefficiency of monopolistic chaebol businesses which constantly demanded increased state support, and, finally, the chaebol-bureaucrat collusion.

Central to such structural problems is not that the state controlled the market for economic development but, rather, that, for some reasons, it lost flexibility to move from control to liberalization when the volume of the economy reached a certain level. Liberalization of the market would not only enhance the competitiveness of chaebol businesses but also allow small- and medium-sized enterprises to grow. As a matter of fact, liberalization was considered by post-Park policy makers. But such considerations either did not materialize in concrete policies or materialized only superficially. More seriously, even the policies of controlling and allocating strategic resources for economic development often did not go through as intended. That is, the developmental state actually did not and could not unilaterally guide economic activity in the direction it desired politically. For this reason, non-statist writers say that "there is little reason to call the rapid economic growth 'government-led'" (Yoo, 1997).

Why did the strong state lose flexibility in policy making and implementation and could not even carry out its policies as it intended them? The "institutionalists" point to the existence in society of "a concrete set of social ties which bind the state to society and provide institutionalized channels for the continual negotiation and renegotiation of goals and policies" (Evans, 1992). Due to such social ties (apparently pervaded by the personalist ethic) and the "institutionalized channels" based thereon—so goes their argument—the state allowed for societal input to its policies and frustration of imposed policies, and this while "dominat[ing] over social constituent" (Lee YH, 1997, p. 8). The structural problems of the South Korean market, thus viewed, were caused essentially by the personalist ethic which hampered both the healthy development of a market economy and the statist state's flexible response to market changes for the purpose of stimulating the voluntary activities of supply and demand. And the ongoing economic crisis means the breakdown of the problematic market structure at a time when the globalizing world market drastically reduced the state's ability to control the nation's economy.

It is here to be noted that, in the institutionalist view, the social channels do not originate solely from such social/personal ties, for imposed state policies, when institutionalized through negotiation and renegotiation with societal forces, often turn against new, more efficient state policies. A writer of the same institutionalist persuasion lays down the process in three phases (Rhee 1994, p. 15). First, there are certain institutional structures—state-imposed or indigenous to society—determining "practices, norms, rules, organizational structures, and organizations that guide both internal relationships within groups and interactions among groups." Once implemented, such structures immediately decide on "individual interests (or preferences) and abilities (or capabilities)." Eventually, their impact becomes far greater because they set limits on further state policies as well, because they "limit the range of policy-making and implementation options available to policy makers." Thus, second, institutional structures, which were initially a variable dependent upon state intervention, turn to be independent variables. As a result, finally, they attain "institutional continuity (or persistence)" which readily overcomes challenges by more efficient policy alternatives. The intervening strong state can thus slip into its own trap, continually exacerbating the structural problems of the economy (and society) it controls.

Implications for the Education Market

What kind of interpretation would these conclusions permit about the surging demand in South Korea's education market? Although supply and demand in education too are market phenomena and, as such, susceptible to the state-market/society relationship, differences between education and other markets should not be overlooked. Before proceeding further to address the question, therefore, it is in order to clarify the differences and, on this basis, characterize South Korea's education market in relation to the intervening state. This will facilitate identifying the South Korean state's education policies on the political dimension, their frustration and impairment at the state-society junction, and, finally, the establishment in the education market of certain institutional structures which disrupt the relationship between supply and demand.

(1) The State and the Education Market

The most striking difference between education and many other markets is that, while the latter are generally free and competition is the rule, the former is usually not so: it receives extensive state intervention in most countries. During the last hundred and fifty years or so, this market has been contracting as the state has been taking over an increasingly large part of it, creating a new public sector. While *private education* in the rest of the market is traded for fees, as it had been so ever since Sumerian times (Krader 1963), *public education* is governed by the principle of universality, that the opportunity of receiving educational services should be made available to all. In Canada, for example, public education is delivered free of charge, and with compulsory measures, to all those in a certain age bracket up to a certain level of schooling. Beyond that level, it is provided to all those who are qualified for and desirous of advanced learning and willing to shoulder some necessary financial burden. Private education too is often free or inexpensive, but it does not and cannot afford to satisfy the universality principle. The underlying assumption of this principle is that, unlike other goods and services, education cannot be left entirely to free and open market transactions (West, 1975 and 1970). State intervention commenced on this assumption. Once commenced, however, it did not stop with the creation of the public sector; it expanded over the private sector to regulate educational activities traded there in order to ensure certain standards and often offered financial support as well. In this sense, private education also bears some public character. Not many other markets evidence such thorough state intervention.

But it is to be noted that the extensive state intervention does not relinquish the education market itself. The practice of selling educational services for fees (and buying them by paying fees) remains in both public and private sectors. Strictly speaking, this practice is absent only when educational services are provided free of charge as in some cases of compulsory education. If compulsory education charges fees, the market practice remains to the extent of the fees, no matter how nominal they are. The consumers in this case are only compelled to buy these educational services at a reduced rate. The market practice is more pronounced in the type of non-compulsory public education where students are selected and fees charged. In that case, the trading of educational services involves fairly active participation by both buyers and sellers. Meanwhile, the educational services offered by private schools go through the market process even when they are subsidised by the state, for state subsidies here only restrict, and do not abolish, the practice of market transactions. The market practice of trading educational services, which remains in many of today's private and public schools, helps us determine the degree and extent of state commitment to education and, thereby, distinguish between the realms of the political state and the market in education. The political economy theses on South Korea's state-society relationship may be looked at in terms of this distinction.

(2) A Statist Political Economy of Education

The situation in the South Korean state's intervention in the education market is reflected in the interesting use of the terms "public education" and "private education" in the country's education literature. In that literature, the former means education provided by any regular grade school or university, while the latter stands for education supplied by small or large neighbourhood cramming classes. Regular private schools—namely, private primary, middle and high schools and universities—are all considered institutions of public education while public schools and state universities, of course, remain so. And since the neighbourhood cramming classes are undesirable, South Korean writers—both professional researchers and occasional commentators on educational issues—argue for the "eradication of private education." In fact, the latter has been a major catchword in state education policies as well since the last years of military rule (surprisingly, in spite of the staunch anti-socialist position).(Note 12) What this uniquely South Korean terminology suggests is the magnitude of state intervention which by virtue of control has transformed *de jure* private schools and universities to *de facto* public institutions.

We are confident of this assertion and base it in part on an examination of South Korea's private schools. These schools indeed appear to be public in the sense that they are subject to the state's extensive control as much as are public schools. They admit students within the quotas set by the state and as allocated by the state (either through direct allocation or through state-administered entrance competition), collect tuition and other fees as set or approved by the state, and teach the same state-imposed curriculum using the same state-made or state-approved textbooks as in public schools. A private school teacher is by law a "civil servant" just as his/her public school counterpart: both are agents carrying out the state's decisions. This legal view is firmly entrenched in state policies, as evidenced by the current policy to "institutionalize teacher rotation between private and public schools."(Note 13) Although universities do not receive direct state intervention in their curriculum, the curricula which they actually employ are fairly uniform because of other state policies (especially the system of state-administered entrance competition as we shall see shortly.)(Note 14) In such spheres as student quotas and tuition fees, they too are subject to rigid state control with no distinction between private and public.

The delivery system of educational services, on the other hand, maintains the market mode of transaction in all schools and universities, private or public. While they are all treated as "institutions of public education," none of them actually provides free educational services. As mentioned earlier, primary schools conducting compulsory education raise more than 10 percent of their expenses from their pupils. Tuition fees of public secondary schools and state universities are only slightly lower than those of their private counterparts. Educational services are invariably traded for fees while the content, volume and price of the services are kept under the state's uniform control.

The origin of these phenomena can be traced to the developmental state's policies. At the time of Park's 1961 coup, the country was still in the ruins of the Korean War, which had ended only eight years before. The state's coffers were depleted. The First Five-Year Economic Development Plan, which Park's regime launched in 1962, required a large portion of the resources under state control for supporting strategic industries. Meanwhile, full participation in primary education in the late 1950s increased pressure on the regime by increasing the demand for education in middle schools and by creating severe classroom and teacher shortages in primary schools. The regime's response was to minimize state expenditures on education and concentrate the secured education budget on primary schools. The regime could employ this strategy in spite of its ambitious economic development plan because manpower requirements from the early stage of industrialization had been basically met by the universalization of primary education (Cho, 1994, p. 101n). In effect, the annual education budget during Park's 18-year rule ranged between 1.9 and 3.0 percent of the GNP and between 14.4 and 18.0 percent of the total state budget.(Note 15) From the education budget, between 70 and 82.8 percent went to primary education, the remainder being split among secondary and tertiary education, as well as the huge education bureaucracy which the statist regime required as an instrument of government (KEDI, 1997b, p. 229).

The secondary and tertiary education policy which Park's regime could work out under these circumstances was mainly to contain expansion in order thus to avoid possible demand for increased state funding (although the regime officially put forward, as reasons, improvement of quality in education) (e.g., Lee IY, 1996). For this, on the one hand, the regime downscaled universities and colleges, restricted licencing of new institutions, controlled student quotas for educational institutions and programs, and intervened in student selection by introducing a state entrance examination (in order to prevent frictions in selective admission of students). On the other hand, it chose to use existing private schools and universities by subjecting them to close supervision and extensive control for the purpose of (partially) meeting the rising demand, instead of the costly option of increasing public secondary schools and state universities.

Tuition fee control—where fees are set by the state rather than by the educational institutions concerned—was more closely related to developmental policies, because it was employed primarily as a measure to fight inflation, important as the latter was for not losing gains from economic development. It also acquired significance as an education policy as entrance competition intensified. Since this competition was basically due to the shortage of supply in student places in secondary and higher education, its intensification aroused public concerns about insufficiency in state funding and heavy financial burdens on parents (e.g., Min, 1963; Chon et al., 1969). The Park regime, in response, more tightly controlled tuition fees and student selection as a way to allaying the discontent. Given that secondary and post-secondary education was not to be available to all but a few, low tuition fees and state-administered student selection would yield a system of meritocracy, in which those who had successfully survived state-administered competition would be somehow able to pay for the hard-earned opportunity for education. (Note 16)

Such military-controlled developmental state's policy measures pushed demand in education higher than it would have been otherwise. So long as the state held down the supply of education, it was natural that entrance competition intensified. In addition, the uniformity of learning ensured by the uniform teaching materials—state-imposed curriculum and textbooks—laid a common ground for the nationwide entrance competition and turned classroom instruction into rote-learning in state-imposed basics and drills, essential as these were for preparation for the entrance competition. The intensifying entrance competition quickly swept students into the movement toward schools of higher level and caused demand in education to rise further. During the 1960s alone, as a result, the quota-controlling Park regime had to moderate its supply control and raise university student quotas by as much as 43 percent, because the number of applicants to universities had nearly doubled (Kim YB, 1980, p. 271). This moderation took place although university entrance competition at that time was much less serious than the competition for middle school entrance.

Since then, the military regime's typical policy regarding entrance competition was to maintain supply control until the intensifying entrance competition and public discontent reached a critical point and, then, to abruptly modify the control at the level of schooling where applicants were most congested. The measures typically employed for this were to increase student quotas and to loosen conditions for licencing new institutions and programs. Such crises first hit middle schools in the late 1960s, high schools in the late 1970s, and universities in the early 1980s. The increase of student quotas, however, invariably ended up further enhancing demand because the intensity of the entrance competition, which had already become critical, drove students toward higher levels of schooling with an uncontrollable force. The demand which was thus enhanced quickly exceeded the increased student quotas.(Note 17) Thus, while the number of schools, universities and programs increased rapidly, demand for education grew even more rapidly. To a large extent, the incredible speed of growth in South Korea's education can be accounted for by such early developmental policies.

Overall, the Park regime's initial decision to control supply in the education market appears to have been inevitable on the political dimension of policy making. When economic development was to be pursued from scratch and necessary manpower for industrialization was already secured by the universalized primary education, expenditures on not so urgent areas, such as education, might well have to be suppressed. The resultant intensification of entrance competition and even the rise in demand could be reasonably tolerated for a while. The trouble, however, was, as in the case of economic policies, the state's inability to move from the temporary measures to policies which were more viable in the long run. As the growing economy gradually improved the condition of its coffers, the state could loosen its grip on the education market and liberalize private educational institutions in terms of their program development, student admission and decisions on tuition fees. At the same time, it could increase investment in the public sector to expand and consolidate its system. Were freedom restored for private institutions to conduct education according to their own decisions, supply and demand in education would sooner or later be balanced, for the consumers in that case would consider which schools and what programs were useful for them, instead of blindly seeking higher and more education. On the other hand, concentration of financial resources on the public sector would not only enhance public education but also motivate private schools' competitive efforts for survival and prosperity via marketable programs. But the state did not choose such a free market alternative. Instead, it continued with supply control and further tightened its grip on private education. And as part of this policy, it committed much of the increased education funds to subsidising private schools (if not private universities).(Note 18) Thus, it continued to prop up the rising demand for more supply in education, contrary to its original policy pronounments.

The failure of the state to make policy changes may not be explained by reference to the political dimension of policy making. It may be more plausibly explained by referring to the implementation process in which state policies met the economic world of the personalist society. Especially notable in this connection are the behaviours of those parties who were involved in the transaction of educational services and of the officials functioning on behalf of the intervening state.

(3) The Personalist Ethic and Protection under State Control

The statist policies of controlling student quotas and tuition fees, and of prohibiting diversity in curriculum and textbooks, were obviously not what schools, especially private schools, would welcome. Quota control and tuition fee control meant constraint on their financial operations. The control of curriculum and textbooks flagrantly violated their fundamental rights to conduct education according to their own philosophy and professional judgment. Nevertheless, when policy makers for the first civilian government of Kim Young Sam contemplated in 1994 "liberalizing" educational institutions, the fiercest opposition came from private institutions themselves. Not so conspicuous yet unambiguous resistance came also from within the state bureaucracy, in spite of the fact that it was the very instrument to carry out state policies. As a result, most liberalizing elements in the government's education reform schedule had vanished by the time it was made public. This irony can be accounted for in terms of the personalist ethic which affected the operation of the parties involved in the supply and demand of educational services.

As regards the private educational institutions, although state control restricted their freedom and violated their rights, it practically permitted them advantages which far outweighed disadvantages on their personalist ethical scale. Most importantly, it guaranteed a secure way of balancing revenues and expenditures. State-set student quotas maintained supply in education constantly short of demand. Meanwhile, the state-administered entrance examination repeatedly invigorated competition among students for entry to higher level schools. In effect, secondary and post-secondary institutions always received more applications than their quotas. Although the state placed limits on the number of students to admit and the amount of tuition fees to collect, the limit was not so low as to bankrupt the private institutions, for the state's policy was to rely on private institutions rather than expand public institutions to cope with the rising demand. The student quotas and tuition fees which were actually granted to private schools and universities, therefore, were usually sufficient to raise necessary funds for running the latter's programs at minimal levels. If private institutions relinquished control over the content of their programs and kept costs of instruction (teacher salaries in particular) at low levels, they could make ends meet and even generate some profit.(Note 19) This way of operating their institutions under state control and with "protection of vested interest" was much easier and safer than trying to draw students by offering attractive programs in an open market.

Many private institutions, in fact, pursued their own interests more actively. Since they mostly relied on tuition fees for their revenues, the key to gainful operation was the securing of as many students as possible. The state's student quota control, however, practically blocked this avenue. Furthermore, the existing student quotas, albeit under control, had quickly surpassed the level of normal classroom operation as repeated crises in entrance competition compelled the state to raise student quotas. Leeway was found in the increase of education programs (that is, more classes for high schools and more departments and faculties for universities) and affiliated institutions (such as elementary and secondary schools for universities and universities for secondary schools). Many private schools and universities succeeded in this kind of effort, especially between 1980 and 1997. Since entrance competition at universities created a continual crisis during this period, the state repeatedly chose to institute restricted yet large-scale increase in the total supply of student places. Taking advantage of the large sums of tuition fee revenues and the multiplied collateral values of their school properties (due to economic development), private schools lobbied and obtained licences for new programs and affiliated institutions as well as necessary favourable bank loans. As a result, many of them grew to be large educational complexes comprising numerous schools at different levels and accommodating tens of thousands of students. At the same time, they incurred large debts which were payable only when cash kept flowing in not only from tuition fees but, more importantly, from additional bank loans. And, as in the case of chaebols, justification of such loans required continual expansion of the private educational institutions. The vicious cycle of expansion for the sake of expansion could be sustained only when entrance competition provoked demand for more and higher education. Hence, ironically, private schools and universities demanded that the state continue to control their operation, keep supply short, and maintain other policy measures which exacerbated entrance competition.

The grounds for collusion between state officials and educational institutions, then, are obvious. While a strong state controlled educational institutions in all aspects of their operation, concrete decisions in dealing with individual institutions were made by governmental leaders and state officials whose *modus operandi* was as personalist as most others in society. The tighter and broader the state exercised control, the greater became the arena for collusion. And through personalist collusion, new school and university licences were given often to those with no backgrounds in the educational profession, and permits for existing institutions' additional programs and affiliated schools were granted without proper assessment of feasibility. A substantial portion of the state's financial resources as well were wasted through collusion.(Note 20)

Another ironic phenomenon in South Korea's education market is the ambivalence of the consumers. Although they were the victims of the inflexible state policies and personalist collusion, they too were reluctant to support changes to the existing state policies. Generally, they understood that schools and universities were operating as businesses rather than as non-profit bodies as legally defined. They also agreed that state control of student quotas and student selection aggravated entrance competition. Their conclusions, however, were that the state should more strictly control schools and universities in the areas of student quotas and tuition fees, and that the state entrance examination should yield scores which more precisely distinguished between individual aspirants for higher education, although doing so should further exacerbate entrance competition. Underpinning this attitude was undoubtedly the personalist ethic. What was of immediate concern was not the development of a viable education system but their own personal interest in not becoming a loser in the nationwide competition for higher and more education and, in case they won, in not paying high tuition fees even at tuition fee-dependent private institutions.

All these parties turned against the reforms intended to accord educational institutions autonomy and freedom necessary for diversifying the education market and, thereby, reducing entrance competition and unnecessary demand for education.(Note 21)

(4) Institutionalization of Statist Policies

The developmental state's educational policy measures, which were initially introduced for the purpose of temporarily containing the expansion of secondary and tertiary education as a way to secure resources for strategic industries, have now become "institutionalized" in the education market although the need for such measures has disappeared. The South Korean state's education budget, which now consumes 5 percent of the world-class GNP, is comparable to those of many advanced nations of similar population. Participation in primary and secondary education is full and in tertiary education it is nearly universal. Even so, the state continues the containment policy, allocates students to schools and universities directly or via state-administered entrance competition, controls tuition fees, and maintains the uniformity of curriculum and textbooks in grade schools. It does so on the statist assumption that private schools as well as universities are institutions of public education. And indeed the South Korean state does not seem to be able to abandon such policies unless it resolves to overhaul the entire education market and risk social and political dislocations.

The main source of such difficulties is obviously the prevailing personalist ethic. The hurdles erected by this ethic are interestingly reflected in the remarks made by the critics of liberalizing policy ideas. For instance, in 1998 a professor of mathematics education at a private university in Seoul criticized the policy proposal to abolish entrance examinations and argued that, should this institution be abolished, the state would have to supply alternative ways of determining whom to accept and whom to reject. The leader of a major parental organization in the same year opposed liberalization of schools—especially the abolition of the student quota system and tuition fee control—for the reason that the "greedy" owners of private schools and universities would immediately raise tuition fees and admit students with no limit.(Note 22) Both leaders of public opinion could not see beyond the statist political economy because they did not consider that, where state control and protection is absent or not as extensive as in South Korea, whom to admit and whom to reject depends on what kind of education the university provides, and that, in the private sector, the tuition fee is one major regulator of the supply-demand relationship. On the other hand, state bureaucrats continually seek to influence policy makers in order to expand the extent of state intervention instead of shrinking it. None of the two civilian governments that have so far emerged since the end of military domination was prepared to confront such obstacles. It was not that they were unwilling to dismantle old institutional structures, but simply that the electorate, divided along personal connections, would not confer on them the clear-cut mandate they needed.

The statist policies, which became indispensable institutions in the state-controlled education market, have now acquired "institutional continuity" and, as such, "readily overcome" challenges by more viable policy ideas. The strong South Korean state has, thus, fallen into its own trap, continually expanding its intervention while seeking liberalization, and exacerbating the question of entrance competition while trying to solve it.

(5) The Demand-Provoking Institutional Structures

This being said, one last question remains to be addressed. Why, as in 1998, does demand continue to rise and university entrance competition remain intense even when the overall supply in student places has been increased far beyond the overall demand? The answer can be obtained by examining an institutional structure that has emerged on the supply side over the years of state intervention. This institutional structure is the hierarchy of universities which South Koreans unanimously accept.(Note 23)

This hierarchy ranks universities not on the basis of their performance in teaching and research, nor on the basis of their educational facilities and equipment, but on the basis of a few external conditions. The conditions are (i) whether a university is located in Seoul or elsewhere, (ii) whether it is a state or private university, and (iii) whether it is older or newer than others.(Note 24) Preference is based on the assumption that the university is better if it is located in Seoul not elsewhere, a state university not a private university, and older not newer than others. Superficial though it may appear, this assumption is not groundless. Rather, it is fairly well grounded in a web of institutionalized state policies. Consumer indifference to the actual performance of individual universities has to do with the state policies of keeping supply short of demand and provoking entrance competition. The reason is that a guaranteed supply of students under perpetual entrance competition deprives the universities of incentives for trying to outperform others or to develop unique education programs. Since all universities, thus, tend to provide educational services in money-saving ways, consumers look at conditions other than their programs, teaching and research, or equipment and facilities, when they consider which of them to choose. Preference for universities in Seoul is explained partially by the established state policy of not licencing any new university in Seoul in order to curb the concentration of population, which in turn was prompted by another state policy of seeking industrialization at the expense of agriculture.(Note 25) The higher tuition fees of private universities for larger classes with less than satisfactory facilities and equipment may account for the low popularity of such universities. Licences issued to new universities at times of crisis are part of the reason for preferring older universities. Finally, the personalist ethic which requires useful personal connections to the state bureaucracy and chaebol businesses endorses the status of the universities ranked high because their graduates are already in control of those institutions. Since these conditions are not what individual universities can alter, the hierarchy remains constant and individual universities' ranks fixed, with the state university in Seoul at the top and the private universities springing up in the country at bottom.

The hierarchy's role in accommodating demand in higher education is obvious in the fact that it is a hierarchy based on judgement about which university is preferable to which. The top-ranking university is the most preferred while the bottom one is the least. Most aspirants for higher education follow this rule in seeking admission. This tendency is so strong that it persists even after entry to first-year programs well into advanced levels of undergraduate studies. Each semester universities invite applications for the vacancies created by those who do not register because filling given quotas is vital for revenue generation. This is immediately followed by the breaking out of transfer competitions which are as intense as entrance competition.(Note 26) Since the upper ranks are always taken by a handful of prestigious universities while all aspirants for higher education seek to go to such universities, the actual gap between supply and demand remains great even when the overall supply surpasses the overall demand and lower-ranked universities fail to fill their quotas.

The significance of the hierarchy's role in provoking entrance competition can be considered in connection with the practices of state entrance examination and the uniform curriculum and textbooks in grade schools. The once-a-year state entrance examination gives all aspirants for higher education scores specified down to decimal fractions to sort them into thousands of minute ranks. Its questions are prepared in a multiple-choice format about the contents of the uniform curriculum and textbooks in selected subjects with minor variations. Since this examination is the only road to higher education while its questions are such that better prepared students perform better, it motivates most students in grade schools to competitive preparation, including those with no serious interest in advanced learning. While schools conduct instruction that focuses on entrance competition, their staying away from the competition means alienation from the mainstream school life. Meanwhile, the hierarchical pyramid of universities lifts the nationwide drive toward higher education up to its apex, the state university in Seoul and a handful of prestigious private universities in the same city. In sum, the hierarchy of universities keeps the preferred student places in short supply while the state entrance examination pushes students toward top-ranking universities, and the uniform curriculum and teaching material makes the procedures of entrance competition a simple and technical matter.

Summary and Conclusion

If the foregoing discussion is not flawed, it may permit a conclusion that the high demand for education in South Korea is a product of the statist political economy which, in spite of its initial "political" commitment to coordinating educational development with state-initiated economic development, has become entangled in the private interests of a society in which a personalist ethic prevails. As part of the developmental strategy, the state initially sought to regulate the growth of secondary and higher education in pace with the growth of resources available for education. It also decided to rely on private education rather than expand public education for meeting the demand. That decision was clearly geared to develop a solid market in education as well as in the economy. The state, nevertheless, could not abandon such initial policy measures even when the changed circumstances no longer required them, and even after it practically dropped "development" from its policy agenda. The four decades of education control have resulted in private educational institutions complacently seeking prosperity under the state's control and protection. Here, the personalist ethic of traditional Korea helped bond the policy measures of control and protection to the personalist ethic of the agents in the education market and the officials of the intervening state.

Thus, even at a time when the overall supply of student places far exceeds the overall demand, the institutionalized policy measures continue to provoke entrance competition and, thereby, demand for education in higher and more prestigious institutions. They do so by freezing student quotas at prestigious universities, by keeping tuition fees low enough to freeze the content of education at low standards, and by selecting students on the basis of scores from state examination in the basics covered in state-imposed curricula and textbooks. All this work together to sweep schools and their students into the nationwide entrance competition, thus producing the rising demand. Meanwhile, the hierarchy of universities maintains the intensity of entrance competition by keeping supply of student places at higher ranking institutions always short and by directing the drive toward higher education to the high-ranking universities.

Undoubtedly, the near universalization of higher education in South Korea is a remarkable feat. But there are serious problems as well. The most serious one is that the selection of university students by competitive entrance examination. Although this method has been a vital factor in the achievement of near universal higher education, it prevents what UNESCO (1998) calls "diversification of higher education" and makes it impossible for higher educational institutions to be "a lifelong source of professional training, updating and recycling." The former is due to the guaranteed supply of students which weakens incentives for developing unique programs; the latter comes from the fact that competitive entrance examinations block the existing workforce from returning to institutions of higher education. Moreover, the fact that almost all South Koreans are about to become university-trained professionals suggests a serious weakness of the future workforce, for a healthy workforce includes not only professionals but also skilled and semi-skilled workers. Such weaknesses will place South Koreans in a very difficult situation given that the globalizing world market requires well-trained, versatile workers who can nimbly adjust to the changing world of work.

The "structural problems" must be addressed soon. The institutionalized statist policies must be changed and ways explored to develop the personalist ethic to a *modus operandi* which is more viable in the new world market. Will lessons learned from the South Korean case not cast light on other East Asian nations where statism and the personalist ethic still prevail?

Notes

1. In 1997, participation rates were 98.6 percent for primary schools, 101.0 percent for middle schools, and 94.6 percent for high schools (KEDI 1997a, 32). Primary schools' participation rate was 96.4 percent in 1959 and stayed above 100 percent for most years since then. Dropout rates in South Korean schools are negligible, e.g., 2.1 percent for high schools in 1995 (Ibid., 126) while that for American high schools was 5.3 percent in 1994 (Fetler 1997).
2. The rate rose very quickly in recent years. It was 60 percent in 1997 and 51 percent in 1995 (KEDI 1997a, 34). The non-mature

students who were admitted in 1998 to day-time university and junior college programs alone accounted for 73 percent of all high school leavers. This is comparable to Japan's 35-38 percent between 1988 and 1998 (Mombusho 1988-1988).

3. The proportion of fees other than tuition fees in the operation costs was 17.9 percent in 1980 and 14.4 percent in 1995 respectively (KEDI 1997a, 59).
4. Regarding the *juku* cramming classes in Japan, Simmons (1990, 90) writes: "[The *juku* phenomenon] is a further measure of Japanese enthusiasm for education, or Japanese anxiety about educational achievement, and it is certainly a sign that the Japanese are willing to pay for educational assistance and advantage."
5. There should be a limit in the Confucian culture's contribution to educational and economic development. Lie (1992, pp. 295-6) points out that the Confucian culture has been around in Korea for centuries yet Korea's rapid economic and other development is a recent phenomenon. Long before, Weber (1967) expressed his skepticism about Confucianism's possible contribution to development.
6. This view is frequently found in South Korea's press.
7. This condition is not met in South Korea. In 1997, out of 2,792,410 students in higher education only 126,075 (4.5 percent) were recipients of a scholarship, of which values were mostly far short of paying necessary expenses (MOE 1997, 578, 592). Student loans are virtually unavailable.
8. For the concepts of "civil society" as a society of capitalist market economy and "the state" as a political entity overseeing civil society, I follow the long-established yet essentially controversial tradition in social science. See Marx (1975).
9. Chang (1991) goes further to say that the development of capitalism does not require the individualist ethic. Favourable interpretations of the personalist ethic in Asian countries are also found in Smith (1991) and De Vos and Hsu (1985).
10. This is a South Korean application of Gerschenkron's thesis that "a developing country tends to establish its industrial structure differently from the advanced countries and . . . the higher the degree of backwardness, the stronger will be the role of the government." Cited from Cho (1995, 7). For Gerschenkron's own account, see: Gerschenkron (1961 and 1989).
11. The Germany-based Transparency International's transparency index (TI) for South Korea is lower than Singapore, Taiwan, Hong Kong and Japan while higher than Indonesia, Thailand and China. TIs for East Asian nations are generally lower than those for Euro-North American nations. See Http://www.gwag.de/~uwvw/cpi1998.html.
12. It was at that time that researchers have started to compound public and private education. The earliest example is Yun and Park (1977).
13. "*Kongsaribkan kyosasunhoanje*" (Teacher Rotation between Public and Private Schools), *The Chosun Ilbo*, 7 Sept. 1998.

14. See below under the heading of "The Personalist Ethic and Protection under State Control." While state-administered entrance competition enhances the drive towards higher education and, thus, guarantees universities' filling student quotas, the universities which have no trouble in filling student quotas do not need to recruit students by attractive education programs. Most universities, therefore, copy the programs of the top-ranking university (less bits of something as theirs do not have to be as good as the latter's).
15. For comparison, the size of the 1998 education budget was 5 percent of the GNP and 24 percent of the total state budget. Given that the economy during the Park years was much smaller, it is obvious that state funding of education was much less than adequate for balanced development in schools of all levels.
16. Later on, the state allowed private schools and universities to "autonomously decide" on their tuition fees. Yet the "autonomously decided" tuition fees had to fall within the state's expectation in order for the schools and universities not to be retaliated against.
17. Choi (1996) pointed out that as a result of the Chun-Roh junta's policies, the number of university entrants increased by 30 percent and reduced young university graduates' wages down to the level of young high school graduates.
18. See Note 20.
19. Up until the late 1980s teachers' salaries were very low while their teaching load was very heavy. In 1989, disgruntled teachers unionized against the law prohibiting such an activity. The military regime discharged those teachers who joined the illegal union and arrested their leaders. At the same time, it decided to increase state funds for teacher salaries. This decision was applied not only to public schools but also to private schools. The purpose of this was to mollify private school teachers' discontent while maintaining the existing arrangement for minimal-cost school operation under state control and protection.
20. An administrator whom I interviewed in May 1998 gave me his estimation that at least a quarter of the state education budget was wasted in corruption.
21. The narrow perspective of the political state and class-dominated civil society is apt to lead to overemphasizing the "political" role of the state while overlooking the dynamics in which state policies protect class and other interests in civil society. An example of this is Chung and Armer (1997), which emphasizes the developmental state's role "in shaping policies to expand and limit education to meet political and economic objectives rather than class interests." In this account, the inability of the strong state to carry out its political decisions is simply ignored.
22. The former was found in a newspaper article and the latter in proceedings of a public forum. I withhold documentation in order to avoid possible injury to the persons concerned.
23. The existence of hierarchy in East Asian universities is partially noted in Hayhoe (1995).

24. The general background description of this hierarchy can be seen in Kwon (1991).
25. As of 1997, for instance, only 39 out 169 universities were located in Seoul although a half of the whole population resided in that region (MOE 1997, pp. 580-9).
26. Each year, more than 7 percent of university students migrate upward in the hierarchy. The number of registered university students in 1997 was 1,368,461. At the beginning of the Second Semester of the 1998 academic year, 42,468 of them moved to another university. Those to be transferred at the beginning of the First Semester of the 1999 academic year are expected to be well over 50,000. "Sasang choedae-yi taehag pyoniphag" (The Largest University Student Migration in History), *The Chosun Ilbo*, 17 January 1999. The transfer application rate at major private universities in Seoul for the First Semester of the 1999 academic year ranged between 12 for 1 and 66 for 1.

References

Amsden, A.H. (1994): "The Specter of Anglo-Saxonization Is Haunting South Korea." In: *Korea's Political Economy: An Institutional Perspective*, edited by Lee-Jay Cho and Yoon Hyung Kim (Boulder: Westview, 1994), 87-126.

Chang, Yun-Shik (1991): "The Personalist Ethic and the Market of Korea," *Comparative Studies in Society and Culture*, 33:1 (1991), 106-29.

Cho, Soon (1994): *The Dynamics of Korean Economic Development* (Washington, DC: Institute for International Economics).

Choi, Kang-Shik (1996): "The Impact of Shifts in Supply of College Graduates: Repercussion of Educational Reform in Korea," *Economics of Education Review*, 15:1, 1-9.

Chon, Nyongi, et al. (1969): "Nyongan tuipdoenun kyoyugbi chujong yongu." Research report, Central Institute for Research in Education.

Chung, Insook, and Armer, J.M. (1997): "State, Class, and Expansion of Education in South Korea: A General Model," *Comparative Education Review*, 38:4, 531-45.

Dahl, R.A. (1967): *Pluralist Democracy in the United States: Conflict and Consent* (Chicago: R. McNally).

De Vos, G., and Hsu, F. (eds.) (1985): *Culture and Self: Asian and Western Perspectives* (New York: Tavistock).

Dore, R.P. (1982): "Groups and Individuals," in *Society in Transition with Special Reference to Korea*, edited by Yun-Shik Chang, Taihwan

Kwon, and P.J. Donaldson (Seoul: Seoul National University Press, 1982), 13-29.

Dore, R.P. (1980): "South Korean Development in Wider Perspective," in *Korea: A Decade of Development*, edited by Yun-Shik Chang (Seoul: Seoul National University Press, 1980), 289-305.

Evans, P.B. (1992): "The State as Problem and Solution: Predation, Embedded Autonomy and Structural Change," in *The Politics of Economic Adjustment*, edited by Stephen Haggard and Robert Kaufman (Princeton: Princeton University Press, 1992), 139-81.

Fetler. M. (1977): "Staffing Up and Dropping Out: Unintended Consequences of High Demand for Teachers," *Education Policy Analysis Archives*, 5:16 (http://epaa.asu.edu).

Gerschenkron, A. (1961): *Economic Backwardness in Historical Perspective* (Cambridge, MA: Harvard University Press).

Gerschenkron, A. (1989): *Bread and Democracy in Germany* (Ithaca: Cornell University Press, 1989).

Hayhoe, R. (1995): "An Asian Multiversity? Comparative Reflections on the Transition to Mass Higher Education in East Asia," *Comparative Education Review*, 39:2, 299-321.

James, E., and Benjamin, G. (1988): *Public Policy and Private Education in Japan* (New York: St. Martin's).

KEDI (1977a), *Hangug-yi kyoyugchipyo, 1997* (Seoul: KEDI).

KEDI (1997b): *Tonggero bon hangug kyoyug-yi paljachui* (Seoul: KEDI).

Kim, Chong Chol, and Lee, Chong Jae (1994): *Kyoyughaengjong-yi liron-gua shilje*, 2nd ed. (Seoul: Kyoyugkoahagsa, 1994).

Kim, Eun Mi (1997): *Big Business, Strong State* (Albany, N.Y.: State University of New York Press).

Kim, Young Chol, et al., "Sagyoyugbi munje-oa taeungbangan." Research report. Korean Economics and Finance of Education Society.

Kim, Young Hoa, et al. (1994): *Hangugin-yi kyoyungnyol yongu* (Seoul: KEDI).

Kim, Yung Bong (1980): "Education and Economic Growth." In: *Human Resources and Social Development in Korea: Essays on the Korean Economy*, IV, edited by Chong Kee Park (Seoul: Korea Development Institute), 234-73.

Kim, In Hoe (1991): "Hangugin-yi kyoyungnyol, ku ho-oa shil," *Taehagkyoyug*, 50, 71-7.

Kong, Un Bae, et al. (1994): "Hangug kyoyugtuja shiltae-oa suingnyul punsog-e koanhan yongu." Research report. Korean Educational Development Institute.

Krader, S.N. (1963): *The Sumerians: Their History, Culture and Character* (Chicago: University of Chicago Press).

Kwon, Kyun (1991): "Pinnagan kyoyug pungjo." In: *Kyoyungnangug-yi haebu*, edited by Bum Mo Chung (Seoul: Nanam), 93-117.

Lee, Il Yong (1996): "Taehagiphagjedo-yi kaehyoge koanhan pyongayongu," *Hanguk kyoyungmunje yonguso Ronmunjip*, 11 (1996), 133-47.

Lee, Sunwha, and Brinton, M.C. (1996): "Elite Education and Social Capital: The Case of South Korea." *Sociology of Education*, 69:3 (1996), 177- 192.

Lee, Yeon-ho (1997): *The State, Society and Big Business in South Korea* (London: Routledge).

Lie, J. (1992): "The Political Economy of South Korean Development," *International Sociology*, 7:3 (1992), 295-6.

Marx, K. (1967): *Early Writings* (New York: Vantage Books).

McAdams, R.P. (1993): *Lessons from Abroad: How Other Countries Educate Their Children* (Lancaster, PA: Technomic Publishing.)

MCE (Ministry of Culture and Education) (1988): *Mungyo sashimnyonsa* (Seoul: MCE).

Min, Yun Gi (1963): "Hagbumo pudam kyoyugbi chosayongu." Research report, Central Institute for Research in Education.

MOE (Ministry of Education) (1997): *Educational Yearbook of 1997* (Seoul: MOE).

MOE (1988): "1998-nyondo chuyo-ommubogo (Briefings to the President on the Ministry's Major Policies for the Year 1998)." Typescript.

Mombusho (Ministry of Letters) (1988-1998): *Wagakuni-no kyoikushisaku* (Tokyo: Okurasho Printing Service).

Pak, Min Sun (1997): "Two Wheels for Lifelong Learning in Korea: Credit Banking and Multimedia Technology." ERIC Clearinghouse

Document, No. JC97048.

Pak, Nam-Ki (1994): "Hangugin-yi kyoyungnyol ihae-rul uihan taeanjog koanjom," *Kyoyughag yongu*, 32:5 (1994), 185-206.

Rhee, Jong Chan (1994): *The State and Industry in South Korea: The Limits of the Authoritarian State* (London: Routledge).

Ryoo, Jai-Kyung, et al. (1995): "Changing Rates of Return to Education over Time: A Korean Case Study," *Economics of Education Review*, 12:1, 71- 80.

Simmons, C. (1990): *Growing Up and Going to School in Japan: Traditions and Trends* (Milton Keynes, PA: Open University Press).

Smith, A.J. (1983): Japanese Society: Tradition, Self and the Social Order (Cambridge: Cambridge University Press).

Smith, D.C. (ed.) (1991): *The Confucian Continuum: Educational Modernization in Taiwan* (New York: Praeger).

Song, B.N. (1990): *The Rise of the Korean Economy* (Oxford: Oxford University Press).

Sorenson, C.W. (1997): "Success and Education in South Korea," *Comparative Education Review*, 38:1, 10-35.

Tu, Wei-ming (ed.) (1997): *Confucian Traditions in East Asian Modernity: Moral Education and Economic Culture in Japan and the Four Mini-Dragons* (Cambridge, Mass.: Harvard University Press).

Unesco (1998): "World Declaration on Higher Education for the Twentieth First Century: Vision and Action." Adopted at the World Conference on Higher Education, Paris, 9 Oct.

Weber, M. (1967): *The Religion of China*, translated by H.H. Gerth and D. Martindale (New York: Free Press).

West, E.G. (1975): *Education and the Industrial Revolution* (New York: Barnes & Noble Books.)

West, E.G. (1970): *Education and the State: A Study in Political Economy* (London: The Institute of Economic Affairs).

Whitehill, A.M. (Ed) (1988): *Doing Business in Korea* (London: Croom Helm).

Yoo, Jungho (1997): "Neoclassical versus Revisionist View of Korean Economic Growth." Harvard Institute of International Development's Development Discussion Paper No. 588.

Yun, Chong Il, et al. (1996): *Hanguk kyoyugjongchaeg-yi tamgu* (Seoul: Kyoyugchulpansa).

Yun, Chong Il, and Park, Chongnyol (1977), "Kyoyugjaejong-yi hyonhoang-goa munje: kyoyugbi pigyoyongu." Research report. KEDI.

About the Author

Ki Su Kim

Faculty of Education
Memorial University of Newfoundland
St. John's, NF, Canada
A1B 3X8

E-mail: kskim@plato.ucs.mun.ca

Ki Su Kim is Associate Professor at Memorial University of Newfoundland. He has published about seventy scholarly papers, two books and several commissioned research reports, all on philosophical and policy issues in education. He has recently completed a policy-developing project addressing South Korea's question of university entrance competition. The project was commissioned by the Ministry of Education, the Republic of Korea, and its recommendations are now being implemented. He is a graduate of Seoul National University.

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs- Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina—Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Rollin Kent (México)
Departamento de Investigación Educativa- DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul- UFRGS
lucemb@orion.ufrgs.br

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search

BEST COPY AVAILABLE

<CHNO>TM521914
<AUTH>Russell, Michael
<TITLE>Testing on Computers: A Follow-up Study Comparing Performance on Computer and on Paper.
<JNL>Education Policy Analysis Archives; v7 n20 1999
<NOTE>Electronic journal available at htpp://epaa.asu.edu/epaa; 53 pages.
<PTYPE>143
<DESC>*Computer Assisted Testing; Followup Studies; Middle Schools; *Performance Factors; *Test Format
<IDEN>*Middle School Students; Open Ended Questions; *Paper and Pencil Tests
<ABST>Examined the effect of taking open-ended tests on computers and on paper for students with different levels of computer skill using items from the Massachusetts Comprehensive Assessment System and the National Assessment of Educational Progress for 287 middle school students. Results suggest a large effect of mode of administration on student performance. (SLD)

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 340 times since June 8, 1999

Education Policy Analysis Archives

Volume 7 Number 20 June 8, 1999 ISSN 1068-2341

A peer-reviewed scholarly electronic journal.
Editor: Gene V Glass Glass@ASU.EDU.
College of Education
Arizona State University,Tempe AZ
85287-2411

Copyright 1999, the EDUCATION POLICY ANALYSIS ARCHIVES. Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

Testing On Computers: A Follow-up Study Comparing Performance On Computer and On Paper

Michael Russell
Boston College

Abstract

Russell and Haney (1997) reported that open-ended test items administered on paper may underestimate the achievement of students accustomed to writing on computers. This study builds on Russell and Haney's work by examining the effect of taking open-ended tests on computers and on paper for students with different levels of computer skill. Using items from the Massachusetts Comprehensive Assessment System (MCAS) and the National Assessment of Educational Progress (NAEP), this study focuses on language arts, science and math tests administered to eighth grade students. In addition, information on students' prior computer use and keyboarding speed was collected. Unlike the previous study that found large effects for open-ended writing and science items, this study reports mixed results. For the science test, performance on computers had a positive group effect. For the two language arts tests, an overall group effect was not found. However, for students whose keyboarding speed is at least 0.5 or one-half of a standard deviation above the mean, performing the language arts test on computer had a moderate positive effect. Conversely, for students whose keyboarding speed was 0.5 standard deviations below the mean, performing the tests on computer had a substantial negative effect. For the math test, performing the test on computer had an overall negative effect, but this effect became less pronounced as keyboarding speed increased. Implications are discussed in terms of testing policies and future research.

Introduction

Recently, Walt Haney and I (Russell & Haney, 1997) reported that open-ended tests administered on paper to students accustomed to working on computer may seriously underestimate students' achievement. Although previous research on multiple-choice items suggests that the mode of administration, that is paper versus computer administration, does not significantly affect the test taker's performance (Bunderson, Inouye & Olsen, 1989), our study suggests that the mode of administration may have an extraordinarily large effect on students' performance on open-ended items.

Focusing on students participating in a project that placed heavy emphasis on computers, the study indicates that approximately 60% of the students in the Advanced Learning Laboratory (ALL School) were performing adequately on writing tests before the project began. Nearly two years after the program was implemented the same writing tests taken on paper indicated that only 30% of the students were writing adequately, an apparent decline of approximately 30% points. Yet, when the same tests were administered on computer (without student access to word processing tools such as spell checking or grammar checking), nearly 70% of students performed adequately. This significant and startling difference also occurred on National Assessment of Educational Progress (NAEP) reading and science items, which required students to respond to open-ended items (similar to items used in the Third International Math and Science Study, the Massachusetts Comprehensive Assessment System and other state level testing programs). The study concludes that for the students in the ALL School, most of whom are accustomed to working on computers, open-ended test questions administered on paper severely underestimated students' achievement.

Although our findings raise questions about the validity of open-ended test results for students accustomed to working on computer but who completed tests on paper, our study had several shortcomings. As we noted, only one extended writing item was used. Furthermore, no information regarding the extent to which students used computers or the proficiency with which students used computers was available. All of the examinees included in the study were accustomed to working on computers. Thus it was not possible to study the mode of administration effect across varied levels of previous computer use. Finally, beyond scores for a set of open-ended items performed by both groups on paper, no other information about prior academic achievement, such as standardized test scores or grades, was considered.

Despite these shortcomings, the results raise important questions about the extent to which scores for open-ended items administered on paper can be used to make inferences about individual students (or their schools) who are accustomed to working on computers. Moreover, if test scores are used to evaluate the effect increased expenditures for computers have on student achievement, the use of open-ended items administered on paper may also undermine the growing emphasis on educational technology.

In this study, I build on our prior work and overcome the shortcomings of our previous study. Specifically, I improved the study design in five ways. First, the sample of students was broadened to cover a range of prior computer experience. Second, information about students' prior use of computers, preference for writing on computer or on paper, and an indicator of students' keyboarding skill was collected. Third, rather than focusing on one extended writing task, several items were administered. Fourth, rather than focusing specifically on writing, open-ended math, language arts, and science items were examined. And fifth, all items included in this study had been used in state or national testing programs and had been validated previously. Thus, for this study I analyze the effect of computer administration across levels of computer use/proficiency using several open-ended items in the areas of language arts, math, and science. Specifically, the following research questions are addressed:

1. Does the effect reported by Russell and Haney (1997) hold across levels of computer use/proficiency?
2. Does the mode of administration effect occur within and across subject areas and if so, is the magnitude of the effect consistent across subject areas?
3. Do students who prefer to write on paper perform better than predicted on open-ended items administered on paper and do students who prefer to write on computer perform better than predicted on open-ended items administered on computer?

Background

For three decades, educational theorists have proposed many ways in which computers might influence education. Although it was not until the 1970's that computers began having a presence in schools, since then the use of computers in education has increased dramatically (Zandvliet & Farragher, 1997). The National Center for Education Statistics reports that the percentage of students in grades 1 to 8 using computers in school more than doubled from 31.5 in 1984 to 68.9 in 1993 (Snyder & Hoffman, 1990; 1994). Similarly, the availability of computers to students in school increased from one computer for every 125 students in 1983 to one computer for every 9 students in 1995 (Glennan & Melmed, 1996). As the number of computers has increased, theories about how computers might benefit students' writing have proliferated. To a lesser extent, some researchers have carried out formal studies to examine whether writing on computer actually leads to better writing. Many of these studies have reported that writing on computers leads to measurable increases in students' motivation to write, the quantity of their work and the number of revisions made. Some of these studies also indicate that writing on computers improved the quality of writing. In a meta-analysis of 32 computer writing studies, Bangert-Drowns (1993) reports that about two-thirds of the studies indicated improved quality for text produced on computer. However, the extent to which writing on computers leads to higher quality writing seems to be related to the type of students examined. Generally, improvements in the quality of writing produced on a computer are found for learning disabled students, early elementary students, low-achieving students and college-aged students. Differences generally are not found for middle school and high school students.

Learning Disabled, Early Elementary Students and College-Aged Students

Although neither Kerchner and Kistinger (1984) nor Sitko and Crealock (1986) included a comparison group in their studies, both noted significant increases in motivation, quantity and quality of work produced by learning disabled students when they began writing on the computer. After teaching learning disabled students strategies for revising opinion essays, MacArthur and Graham (1987) reported gains in the number of revisions made on computer and the proportion of those revisions that affected the meaning of the passage. They also noted that essays produced on computer were longer and of higher quality. In a separate study, MacArthur again reported that when writing on a computer, learning disabled students tended to write and revise more (1988). At the first grade level, Phoenix and Hannan (1984) report similar differences in the quality of writing produced on computer.

Williamson and Pence (1989) found that the quality of writing produced by college freshman increased when produced on computer. Also focusing on college age students, Robinson-Stavely and Cooper (1990) report that sentence length and complexity increased when a group of remedial students produced text on the computer. Hass and Hayes (1986a) also found that experienced writers produced papers of greater length and quality on computer as compared to those who created them on paper.

Middle and High School Students

In a study of non-learning disabled middle school students, Dauite (1986) reported that although writing performed on the computer was longer and contained fewer mechanical errors, the overall quality of the writing was not better than that generated on paper. In a similar study, Vacc (1987) found that students who worked on the computer spent more time writing, wrote more and revised more, but that holistic ratings of the quality of their writing did not differ from text produced with paper-and-pencil.

At the middle school level, Grejda (1992) did not find any difference in the quality of text produced on the two mediums. Although Etchison (1989) found that text produced on computer tended to be longer, there was no noticeable difference in quality. Nichols (1996) also found that text produced on computer by sixth graders tended to be longer, but was not any better in quality than text produced on paper. Yet, for a group of eighth grade students, Owston (1991) found that compositions created on computer were rated significantly higher than those produced on paper.

Focusing on high school freshman, Kurth (1987) reports that there was no significant difference in the length of text produced on computer or on paper. Hawisher (1986) and Hawisher and Fortune (1989) also found no significant differences in the quality of writing produced by teenagers on paper and on computer. Hannafin and Dalton (1987) also found that for high achieving students, writing on computer did not lead to better quality writing. But for low-achieving students, texts produced on the computer were of a higher quality than those produced on paper.

Summary of Studies

The research summarized above suggests many ways in which writing on computer may help students produce better work. Most formal studies report that when students write on computer they tend to produce more text and make more revisions. Studies that compare student work produced on computer with work produced on paper find that for some groups of students, writing on computer also has a positive effect on the quality of student writing. This positive effect is strongest for students with learning disabilities, early elementary- aged students and college-aged students. All of the studies described above focus on student work produced in class under un-timed conditions. These studies also focus on work typically produced for English or Language Arts class, such as short stories or essays. However, the study presented here focuses on writing produced under formal timed testing conditions in three subject areas, namely language arts, math and science. Specifically, this study addresses the extent to which producing open-ended responses on computer or on paper effects students' performance, particularly for students with different levels of computer use.

Study Design

To better understand whether open-ended test items administered on paper underestimate the achievement of students accustomed to working on computers, six open-ended math, six science, and six language arts items were converted to a computer format and then administered in two modes, paper and computer. In addition, all students completed a computer use survey and performed a short keyboarding test. Finally, an indicator of prior achievement, namely Grade 7 Stanford Achievement Test version 9 (SAT 9) scores, was collected for each student. As is explained in detail below, the indicator of achievement was used to stratify and randomly assign representative sample groups and is used as a covariate for some analyses.

The study focuses on three subject areas: math, language arts, and science. For each subject area, a total of six open-ended items were administered. To decrease the amount of testing time required for each student, students were divided into four groups. Two of these four groups performed the six science items and three of the language arts items, only. For ease of reference I call these groups of students SL and LS. The remaining two groups of students performed the six math items and the other three language arts items, only. These groups of students are referred to as ML and LM. All students completed the computer use survey and performed the keyboarding test. In addition, an indicator of prior achievement was collected for each student.

The study occurred in three stages. During stage 1, SAT 9 scores were collected for each student. In total, four SAT 9 scores were collected for each student: Comprehensive Normal Curve Equivalent (NCE), Math NCE, Language Arts NCE and Science NCE. Once collected, the Comprehensive NCE was used to stratify and randomly assign four groups of students. Two of these groups formed the SL and LS students while the remaining two groups formed the ML and LM students.

During stage 2, all students completed the computer use survey and performed the keyboarding test. During stage 3, a crossed design was used to administer the open-ended items to each group. In this crossed design, the SL students first performed the science items on computer and then performed three language arts items on paper. The LS students first performed the three language arts items on computer and then performed the science items on paper. Similarly, the ML students first performed the math items on computer and then performed the three language arts items on paper. Finally, the LM students first performed the language arts items on computer and then performed the math items on paper. Below, the instruments, sampling method and scoring method are discussed in greater detail.

Instruments

The instruments used in this study fall into three categories: indicators of prior achievement; computer experience; and open-ended tests.

Indicator of Prior Achievement

As described in greater detail below, an indicator of prior achievement was used to assign students to experimental groups and as a covariate during analyses. Since the sample of students was limited to students in grade eight, the students' grade 7 SAT 9 NCE scores were used as the indicator of prior achievement.

Computer Experience

Two instruments were used to estimate students' level of computer experience. First, a survey that focused on prior computer use was administered to all students. Second, all students completed a brief keyboarding test administered on computer.

Student Questionnaire

The survey was designed to collect information about how much experience students had working with computers and, in particular, how they used computers during their writing process. The survey included questions that asked:

1. how long the student has had a computer in his/her home:
2. how many years they have used a computer;
3. how often they currently use a computer in school and at home;
4. how often they use a computer during different stages of their writing process (e.g., brainstorming, outlining, composing a first draft, editing, writing the final draft); and
5. whether they prefer to write papers on paper or on computer.

In addition, the survey asked students to draw a picture of their writing process and to then describe what they had drawn. The purpose of the drawing prompt was to collect information about if and when computers enter the student's writing process. Finally, the student questionnaire asked students to indicate their gender and their race/ethnicity.

To code student drawings, the following guide was used:

- 0 - No computer visible
- 1 - Computer used for final draft only
- 2 - Computer used prior to creating the final draft

When coding drawings, both the drawing and the student's description of their drawing were reviewed prior to assigning a score. All drawings were coded by one rater. However, to examine inter-rater reliability, a sample of 20 drawings was coded by a second rater. For these 20 drawings, there was no discrepancy between the two raters' scores.

Keyboarding Test

To measure keyboarding skills, all students performed a computer based keyboarding test. The keyboarding test contained two passages which students had two minutes apiece to type verbatim into the computer. Words per minute unadjusted for errors was averaged across the two passages and was used to estimate students' keyboarding speed. Both keyboarding passages were taken directly from encyclopedia articles to assure that the reading level was not too difficult.

Although there is considerable debate about how to quantify keyboarding ability (see West, 1968, 1983; Russon & Wanous, 1973; Arnold, et al, 1997; and Robinson, et al, 1979), for the purposes of this study, students average words per minute (WPM) uncorrected for errors was recorded. In each of the scoring guidelines used to rate student responses to the open-ended test items, spelling was not explicitly listed as a criterion raters should consider when scoring student responses. For this reason, students keyboarding errors did not seem to be directly relevant to this study.

Open-Ended Tests

This study examines the mode of administration effect on student performance in three subject areas: science, math, and language arts. To restrict testing time to 60 minutes per test, 6 science items, 6 math items and two sets of 3 language arts items were administered. All items included in this study were taken directly from open-ended test instruments used previously. Sources for items include the National Assessment of Educational Progress (NAEP) and the Massachusetts Comprehensive Assessment System (MCAS).

Language Arts Items

In total, six language arts items were used in this study. Three of the language arts items were taken from the 1999 Spring administration of MCAS. Two of the items were taken directly from the 1988 Grade 8 NAEP Writing Assessment. And the final language arts item was taken from the 1992 Grade 8 NAEP Writing Assessment.

The three MCAS language arts items focus on reading comprehension. For each of these items, students read a brief passage and then answers an open-ended question about the passage. The passages include a poem titled "The Caged Bird", a speech titled "Sojourner Truth's Speech From the 1850s", and a short story titled "The Lion's Share".

The three NAEP language arts items focus on writing. The first writing item asks students to create a narrative piece that describes an embarrassing experience they have had. The second writing item prompt focuses on creative writing and asks students to write a good, scary ghost story. The final writing item tests students' expository writing skills and asks students to write about their favorite story, telling why they like it and what it means to them.

When selecting the items, two criteria were used. First, the time required to respond to the item could not exceed 30 minutes. Second, the amount of reading (if any) students had to complete before responding to the item could not exceed 1 page. The reason for this second criterion was to maximize the amount of time students spent actually responding to the item. It should be noted that all three MCAS items required students to read a short body of text before responding to a question while none of the NAEP items required students to read any text. For this reason, the MCAS items can be classified as primarily measuring reading comprehension and the NAEP items can be classified as measuring writing ability.

After the six items were selected, they were placed into one of two booklets. Two MCAS items and the 1992 NAEP item formed the test booklet titled Language Arts 1. The remaining MCAS item and the two 1988 NAEP items formed the second test booklet titled Language Arts 2.

Mathematics

The mathematics test booklet contained six items. Three of the items were taken from the 1998 grade 8 spring MCAS test and three items were taken from the 1996 grade 8 NAEP Assessment. Two of the math items tested fractions and proportions. Two items focused on students' ability to read and interpret a graph. One item tested students' ability to calculate and interpret means and medians. And the final item focused on students' problem solving skills.

When selecting mathematics items, two criteria were applied. First, the item had to require students to generate an extended (a minimum of one sentence) written response. Second, the item could not require students to draw a picture, diagram or graph. The first criterion was used to assure that students had to compose text in order to perform well on the item. The second criterion was used to prevent students working on computer from having to access drawing or graphing programs.

Science

Like the mathematics items, three of the science items came from the 1998 grade 8 spring MCAS test and three items came from the 1996 grade 8 NAEP assessment. Similarly, all of the items required students to generate a substantial amount of text (more than a sentence) in order to succeed and none of the items required students to draw pictures or graphs. Two of the items tested students understanding of the physical sciences. Two items focused on human biology. One item tested students understanding of electricity. And the final item tested students' ability to design an experiment.

Scoring Criteria

For all of the items, the scoring criteria developed by MCAS or NAEP were used.

All of the MCAS scoring guidelines used a scale that ranged from 0 to 4. For the MCAS items, a score of 0 indicated that the item was left blank or that the student's response was completely incorrect. Scores of 1 to 4 represented increasingly higher levels of performance.

The scales for the NAEP scoring guidelines varied from 1-3, 1-4, 1-5, and 1-6. A code of 9 was awarded to items that were left blank. For all items, a 1 indicated that the student's response was completely incorrect. Scores of 2 to 6 represented increasingly higher levels of performance. To make the scores for the MCAS and the NAEP items more comparable, all blank responses were re-coded as a zero. The resulting NAEP scales ranged from 0-3, 0-4, 0-5, or 0-6.

Converting Paper Versions to Computer Versions

Before the tests could be administered on computer, the paper versions were converted to a computerized format. Several studies suggest that slight changes in the appearance of an item can affect performance on that item. Something as simple as changing the font in which a question is written, the order items are presented, or the order of response options can affect performance on that item (Beaton & Zwick, 1990; Cizek, 1991). Other studies have shown that people become more fatigued when reading text on a computer screen than when they read the same text on paper (Mourant, Lakshmanan & Chantadisai, 1981). One study (Haas & Hayes, 1986b) found that when dealing with passages that covered more than one page, computer administration yielded lower scores than paper-and-pencil administration, apparently due to the difficulty of reading extended text on screen. Clearly, by converting items from paper to computer, the appearance of items is altered.

To minimize such effects, students taking a test on computer were given a hard copy of the test booklet. The only difference between the hard copy of the test booklets received by students taking the test on computer and the original paper version was that the blank lines on which students recorded their responses were replaced by instructions to write answers on the computer.

Prior to beginning a test, students in the computer group launched a computer program that performed four tasks. First, the program prompted students to record their name and identification number. Second, the program presented the same directions that appeared in their hard copy. Third, the program allowed students to navigate between text boxes in which they recorded their responses to the open-ended questions presented in their test booklet. Finally, after a student completed the test, the program presented two questions about the taking the test on computer (described more fully below).

To assist students in recording their responses in the proper text box, the program placed the question number and accompanying prompt above each text box. To help avoid confusion, only one text box appeared on the screen at a time. To move between text boxes, two buttons appeared on the bottom of the screen. The button labeled "Next" allowed students to navigate to the text box for the next question and the button labeled "Back" allowed students to move to the previous question. Below the last text box, a button labeled "I'm Finished" appeared. Once students felt they were finished with the test, they clicked on the "I'm Finished" button. To assure that they were in fact finished, students were asked again if they were done. If so, they clicked the "I'm Finished" button again. Otherwise, they clicked the "Back" button to continue working on their responses.

When students were finished taking the test and had selected the "I'm Finished" button twice, they were prompted with two questions about the test. The first asked students: "Do you think you would have done better on this test if you took it on paper. Why?" The second question asked: "Besides not knowing the answer to a question, what problems did you have while taking this test on computer?" Students were required to answer these questions before they could quit the program.

To create a computerized version of the test booklets, the following steps were taken:

1. An appropriate authoring tool, namely Macromedia Director, was selected to create software that would allow students to navigate between questions and to write data to a text file.
2. A data file was created to store student input, including name, ID number, and responses to each item.
3. A prototype of each test was created, integrating the text and database into a seamless application. As described earlier, navigational buttons were placed along the lower edge of the screen. In addition, a "cover" page was created in which students entered their name and id numbers.
4. The prototype was tested on a class of ninth grade students to assure that all navigational buttons functioned properly, that data was stored accurately, and that items were easy to read.
5. Finally, the prototype was revised as needed and the final versions of the computer tests were installed on twenty computers in the ALL School and twenty- four computers in the Sullivan Middle School.

For all questions, examinees used a keyboard to type their answers into text boxes that appeared on the screen. To enable students to write as much as they desired, scrolling text boxes were used for all items. Although they could edit using the keyboard and mouse, examinees did not have access to word processing tools such as spell-checker or grammar-checker.

Sampling Method

The sample of students was drawn from two Worcester Public Middle Schools, namely The Advanced Learning Laboratory (ALL School) and the Sullivan Middle School. Since the analyses focus on how the mode of administration effect varies across achievement levels and, more importantly, computer use/proficiency levels, the population of students was pooled across the two schools. Before sampling began, a list of all grade eight students in the ALL School and all grade eight students from one team in the Sullivan Middle School was generated. In total, this yielded 327 students. For each student on the list, an indicator of prior achievement, namely grade 7 SAT 9 scores, was collected. Since some of the students were new to the district or had not taken the SAT 9 the previous year, SAT 9 scores were only available for 287 students. Using a stratified random assignment procedure, students were then assigned to one of two groups. Group 1 was then assigned to the Language Arts 1 and Math tests and group 2 was assigned to the Science and Language Arts 2 tests. For each group, this process was repeated again, this time assigning half of the students in group 1 to take the Language Arts 1 test on computer first and the remaining half to take the Math test on computer first. Similarly, half of the second group of students was assigned to take the Language Arts 2 test on computer first and the remaining half took the Science test on computer first.

Those students for whom SAT 9 scores were not available were randomly assigned to one of the four groups. Although their scores are not included in the analyses below, their responses were used to train raters prior to scoring the test booklets for students included in the analyses.

Due to absences and refusals to perform one or more instruments, complete data records were available for 229 students. To be clear, a complete data record was defined as one containing a student's SAT 9 scores, their responses to the student questionnaire, the results of the keyboarding test, and results from at least one of the open-ended tests.

Scoring

To reduce the influence hand writing has on raters' scores (Powers, Fowles, Farnum & Ramsey, 1994), all responses to the open-ended items administered on paper were transcribed verbatim into computer text. The transcribed responses were randomly intermixed with the computer responses. All student responses were formatted with the same font, font size, line spacing and line width. In this way, the influence mode of response might have on the scoring process was eliminated.

Scoring guidelines designed for each item were used to score student responses. To increase the accuracy of the resulting scores, all responses were double-scored. When discrepancies between raters' scores arose, an adjudicator awarded the final score. At the conclusion of the scoring process, one score was recorded for each student response.

To estimate inter-rater reliability, the original scores from both raters were used. The resulting scores were compared both via correlation and percent agreement methods. Table 1 shows that for most items the correlation between the two raters' scores was above .8 and for many items the correlation was above .9. For two of the items on the first language arts test, however, correlations were closer to .7. Nonetheless, this represents an adequate level of inter-rater reliability.

Table 1
Inter-rater Reliability for Open-Ended Items

	Correlation	% Exact Agreement	% Within 1 Point
Language Arts 1			
Item 1	.80	.68	1.00
Item 2	.74	.50	1.00
Item 3	.72	.59	.95
Language Arts 2			
Item 1	.95	.84	1.00
Item 2	.94	.88	1.00
Item 3	.91	.76	1.00
Math			
Item 1	.91	.60	1.00
Item 2	.83	.65	.90
Item 3	.88	.80	.95
Item 4	.94	.80	1.00
Item 5	.84	.90	1.00
Item 6	.70	.75	.95
Item 1	.80	.64	.95
Item 2	.86	.73	1.00
Item 3	.88	.82	1.00
Item 4	.88	.86	1.00
Item 5	.92	.82	1.00
Item 6	.85	.73	1.00

To estimate intra-rater reliability, one rater double-scored 20% of the responses. The resulting scores for this rater were compared via correlation and percent agreement methods. Table 2 shows high correlations between the two sets of scores. Moreover, where discrepancies occurred, the difference between the two scores was never more than one point.

Table 2
Intra-rater Reliability for Open-Ended Items

	Correlation	% Exact Agreement	% Within 1 Point
Language Arts 1			
Item 1	.92	.86	1.00
Item 2	.95	.95	1.00
Item 3	.88	.77	1.00
Language Arts 2			
Item 1	.91	.82	1.00
Item 2	.93	.91	1.00
Item 3	.94	.86	1.00
Math			
Item 1	.92	.77	1.00
Item 2	.97	.91	1.00
Item 3	.98	.95	1.00
Item 4	.96	.82	1.00
Item 5	.88	.91	1.00
Item 6	.84	.82	1.00
Science			
Item 1	.94	.86	1.00
Item 2	.96	.91	1.00
Item 3	.93	.91	1.00
Item 4	.92	.91	1.00
Item 5	.98	.95	1.00
Item 6	.94	.86	1.00

Note that the adjudicated scores were produced for all students and that the adjudicated scores were used for all analyses described below.

Results

This study explores the relationships between prior computer use and performance on four open-ended test booklets. To examine this relationship, three types of analyses were performed. First, independent samples t-tests were employed to compare group performance. Second, total group regression analyses were performed to estimate the mode of administration effect controlling for differences in prior achievement.

And third, sub-group regression analyses were performed to examine the group effect at different levels of keyboarding speed. However, before the results of these analyses are described, summary statistics are presented.

Summary Statistics

Summary statistics are presented for each of the instruments included in this study. The raw data are also available from this point. These original data are presented by the author for othes who may wish to perform secondary data analyses; anyone publishing analyses of these data should cite this article as the original source. For the student questionnaire, keyboarding test, and the SAT 9 scores, summary statistics are based on all 229 students included in the study. For the language arts, math and science open-ended tests, summary statistics are based on the sub-set of students that performed each test. When between group analyses are presented, summary statistics for select variables are presented for each sub-set of students that performed a given test.

Keyboarding Test

The keyboarding test contained two passages. Table 3 shows that the mean number of words typed for passage 1 and passage 2 was 31.2 and 35.0, respectively. As described above, the number of words typed for each passage was summed and divided by 4 to yield the number of words typed per minute for each student. Across all 229 students included in this study, the mean WPM was 16.5. Considering that the minimum WPM required by most employers when hiring a secretary is at least 40, an average of 16.5 WPM suggests that most students included in this study were novice keyboarders.

Table 3
Summary Statistics for the Keyboarding Test

N=229	Mean	Std Dev	Min	Max
Passage 1	31.2	11.2	5	71
Passage 2	35.0	10.6	9	80
WPM	16.5	5.3	4.8	37.8

Student Questionnaire

The student questionnaire contained 11 questions. The maximum score for the Survey was 46 and the minimum score was 10. The scale for each item varied from 0 to 2, 1 to 2, 1 to 3 and 1 to 6. To aid in interpreting the summary statistics presented in table 4, the scale for each item is also listed. In addition to the Survey total score, summary statistics are presented for the Comp-Writing sub-score.

Although comparative data is not available, Table 4 suggests that on average students included in this study do not have a great deal of experience working with computers. The average student reports using a computer for between two and three years, having had a computer in the home for less than a year, and using a computer in school and in their home less than 1-2 hours a week. Furthermore, most students report that they do not use a computer when brainstorming, creating an outline or writing a first draft. Slightly more students report using a computer to edit the first draft. Most students, however, report using a computer at least sometimes to write the final draft. Similarly, most students indicate that if given the choice, they would prefer to write a paper on computer than on paper. Yet, when asked to draw a picture of their writing process, less than half the students included a computer in their drawing.

Again, the divergence between students' preference and their reported use of a computer in the writing process may indicate that when recording their preference some students provided a socially desirable response. If students did provide socially desirable responses, estimating the effect preference had on students' performance will be less precise.

Table 4
Summary Statistics for the Student Quesitonnaire

N=229	Scale	Mean	Std Dev	Min	Max
Years having computer at home	1-6	2.81	1.86	1	6
Years using computer	1-6	4.75	1.51	1	6
Use computer in school	1-6	2.74	1.10	1	6
Use computer at home	1-6	2.75	1	6	
Brainstorm with computer	1-3	1.35	.55	1	3
Outline with computer	1-3	1.50	.57	1	3
First draft with computer	1-3	1.72	.75	1	3
Edit with computer	1-3	1.85	.73	1	3
Final draft with computer	1-3	2.50	.65	1	3
Preference	1-2	1.80	.40	1	2
Computer in drawing	0-2	.59	.72	0	2
Survey	10-43	24.37	5.68	12	41
Comp-Writing	5-17	9.51	2.77	5	17

Indicator of Prior Achievement

Four indicators of prior achievement were collected prior to the study.

Specifically, SAT 9 Composite, Reading, Math and Science NCE scores were collected for each student. Note that the NCE scores provided for this study had been multiplied by 10 before they were supplied by the district office. Thus, the range for the NCE scores was 10 to 990 with a mean of 500 and a standard deviation of approximately 210 (see Crocker and Algina, 1986 for a fuller description of NCE scores). Table 5 displays the mean and standard deviation for each SAT 9 score. The mean score for each subject area and for the composite score for students included in this study is approximately .5 standard deviations below the national average. However, within this sample of students there is a substantial variation.

Table 5
Summary Statistics for Indicators of Prior Achievement

N=229	Mean	Std Dev	Min	Max
Composite NCE	402	150.2	119	888
Reading NCE	391	178	10	990
Math NCE	389	174	10	990
Science NCE	434	172	67	896

Open-Ended Tests

Four open-ended tests were administered in three subject areas: math, science and language arts. As is described more fully above, two versions of the language arts test were administered. Each of the tests was administered to a sample of students. The number of students who performed each test ranged from a high of 117 for Language Arts 1 to a low of 100 for Language Arts 2. Within each sample, approximately half of the students performed the test on computer and half of the students performed the test on paper.

The summary statistics for the total sample of students who performed each test are presented in tables 6 through 9. Since each test contained some items from NAEP and some items from MCAS, it is not possible to directly compare the total test scores to the performance of students in other settings. However, to aid in interpreting the test scores, summary statistics are presented for each item along with the national or state average performance for each item. Note that comparison data for the MCAS items represents the mean score on a 0-4 point scale for all students in the state. Comparison data for the NAEP items represents the percentage of students nationally performing adequately or better on the item.

Table 6 presents the summary statistics for Language Arts 1 and Table 7 presents the summary statistics for Language Arts 2. For all items but one item included in the language arts tests, a score below 3 indicates inadequate performance. For item 2 on language arts test 1, a score below 4 indicates inadequate performance. For all items, many students failed to perform adequately. For all items, the mean performance was below 3 and for four items the mean performance was below 2. This low level of performance suggests these items were difficult for these samples of students.

Table 6
Summary Statistics for Language Arts 1

N=117	Scale	Mean	Std Dev	% Adequate*	Mean on MCAS	% Adequate on NAEP
Item 1	0-4	1.42	1.04	19	1.99	NA
Item 2	0-4	1.50	1.05	16	1.73	NA
Item 3	0-6	2.91	1.41	31	NA	29
Total	0-14	5.84	2.85			

* For items 1 and 2, a score of 3 or higher was considered adequate performance.
For item 3, a score of 4 or higher was considered adequate performance.

Table 7
Summary Statistics for Language Arts 2

N=100	Scale	Mean	Std Dev	% Adequate*	Mean on MCAS	% Adequate on NAEP
Item 1	0-4	1.23	0.98	10	1.74	NA
Item 2	0-4	1.67	1.07	31	NA	51
Item 3	0-4	2.12	0.81	22	NA	25
Total	0-12	5.02	2.29			

* For all three items, a score of 3 or higher was considered adequate performance.

Table 8 displays the summary statistics for the Math test. Again, for all items except number 4, a score below 3 indicates inadequate performance. For item 4, a score below 4 indicates inadequate performance. For all items, the mean performance for this sample of students indicates that on average students performed below the adequate level.

Table 8

Summary Statistics for Math

N=110	Scale	Mean	Std Dev	% Adequate*	Mean on MCAS	% Adequate on NAEP
Item 1	0-4	1.18	1.17	15	2.05	NA
Item 2	0-4	1.26	1.23	18	1.83	NA
Item 3	0-4	1.64	1.03	22	1.84	NA
Item 4	0-5	2.45	1.28	33	NA	28
Item 5	0-3	1.44	.53	2	NA	11
Item 6	0-4	1.99	.89	30	NA	26
Total	0-24	9.96	4.18			

* For items 1, 2, 3 and 6, a score of 3 or higher was considered adequate.
For item 4, a score of 4 or higher was considered adequate.
For item 5, a score of 3 was considered adequate.

Table 9 displays the summary statistics for the Science test. For all items, a score below 3 indicates inadequate performance. For all items, the mean performance was below the adequate level.

Table 9
Summary Statistics for Science

N=102	Scale	Mean	Std Dev	% Adequate*	Mean on MCAS	% Adequate on NAEP
Item 1	0-4	1.89	1.02	32	1.49	NA
Item 2	0-4	1.71	1.21	26	1.70	NA
Item 3	0-3	1.50	.75	8	NA	19
Item 4	0-3	1.21	.67	3	NA	9
Item 5	0-3	1.78	.90	22	NA	52
Item 6	0-4	1.57	1.09	20	1.81	NA
Total	0-24	9.66	4.14			

* For all items, a score of 3 or higher was considered adequate performance.

Clearly, students had difficulty with all four of these tests. For all MCAS items, this sample of students performed at a level below that of other students in the state of Massachusetts. For the NAEP items, students performed about as well or worse than other students in the nation.

Comparing Performance on Computer and on Paper

For each test, approximately one half of the sample of students was randomly assigned to perform the test on computer while the other half performed the test on paper. Tables 10 through 13 present the results of between group comparisons for each test. For each test, an independent samples t-test (assuming equal variances for the two samples and hence using a pooled variance estimate) was performed for the total test score. The null hypothesis for each of these tests was that the mean performance of the computer and the paper groups did not differ from each other. Thus, these analyses test whether performance on computer had a statistically significant effect on students' test scores.

To examine whether prior achievement, computer use or keyboarding skills differed between the two groups of students who performed each test, independent samples t-tests were also performed for students' SAT 9 Composite score, the corresponding SAT 9 sub-test score, Survey, Comp-Writing and WPM. The results of these tests are also presented in tables 10 through 13.

Table 10 shows that on average students who performed the first language arts test on paper performed the same as students who performed the test on computer. Similarly, differences between the two groups' SAT 9 Comprehensive scores, SAT 9 Reading scores, Survey scores, and WPM were not statistically significant. However, Table 10 shows that the mean Comp-Writing score for students who performed the language arts 1 test on computer was larger than the mean for students who performed the test on computer.

Table 10
Between Group Comparisons for Language Arts 1

Paper N = 57 Computer N = 60	Mean	Std Dev	SE of Mean	t-value	Sig.
LA 1					
Paper	5.84	2.65	.35		
Computer	5.83	3.04	.39	.02	.99
SAT 9 Comp.					
Paper	379	145	19		
Computer	421	159	21	-1.49	.14
SAT 9 Reading					
Paper	360	168	22		
Computer	426	189	24	-1.98	.05
Survey					
Paper	24.6	5.6	.75		
Computer	23.9	5.9	.76	.67	.51
Comp-Writing					
Paper	10.1	2.8	.37		
Computer	9.0	2.6	.33	2.09	.04*
WPM					
Paper	17.5	4.3	.56		
Computer	16.4	5.3	.68	1.17	.24

*Significant at the .05 level.

On the second language arts test, table 11 shows that Comp-Writing was the only measure on which the two groups differed. However, for the second language arts test, students who performed the test on computer had higher Comp-Writing scores on average than did those students who performed the test on paper. For all other instruments, the two groups did not differ significantly.

Table 11
Between Group Comparisons for Language Arts 2

Paper N = 45 Computer N = 55	Mean	Std Dev	SE of Mean	t-value	Sig.
LA 2					
Paper	5.07	1.70	.25		
Computer	4.98	2.70	.36	.18	.86
SAT 9 Comp.					
Paper	413	138	20		
Computer	393	146	19	.59	.55
SAT 9 Reading					
Paper	402	173	26		
Computer	376	172	23	.74	.46
Survey					
Paper	24.4	5.8	.87		
Computer	25.1	5.7	.76	-.63	.53
Comp-Writing					
Paper	8.9	3.1	.46		
Computer	10.1	2.6	.36	-2.09	.04*
WPM					
Paper	15.7	4.9	.73		
Computer	17.2	6.5	.88	-1.23	.22

*Significant at the .05 level.

With a few exceptions, the students who performed the first language arts test also performed the math test. However, those students who performed the first language arts test on computer performed the math test on paper and vice versa. For this reason, table 12 indicates that the mean Comp-Writing score for the computer group was higher than that of the paper group. Again, this difference is statistically significant. For all other instruments, Table 12 indicates that differences between the two groups' scores were not statistically significant.

Table 12

Between Group Comparisons for Math

Paper N = 54 Computer N = 56	Mean	Std Dev	SE of Mean	t-value	Sig.
Math					
Paper	10.70	4.34	.59		
Computer	9.25	3.90	.52	1.84	.07
SAT 9 Comp.					
Paper	414	155	21		
Computer	407	154	21	.23	.82
SAT 9 Math					
Paper	401	179	24		
Computer	406	190	25	-.16	.87
Survey					
Paper	23.6	5.98	.81		
Computer	25.1	5.72	.76	-1.29	.20
Comp-Writing					
Paper	9.0	2.66	.36		
Computer	10.1	2.65	.35	-2.22	.03*
WPM					
Paper	16.0	4.7	.64		
Computer	17.9	5.2	.69	-2.01	.05

*Significant at the .05 level.

The open-ended science test was the only test for which there was a statistically significant difference in the two groups' test performance. Table 13 shows that on average the computer group performed better than the paper group. There were no other statistically significant differences between the two groups.

Table 13
Between Group Comparisons for Science

(

Paper N = 51 Computer N = 51	Mean	Std Dev	SE of Mean	t-value	Sig.
Science					
Paper	8.55	3.88	.54		
Computer	10.76	4.14	.58	-2.79	.006*
SAT 9 Comp.					
Paper	388	134	19		
Computer	426	152	21	-1.33	.19
SAT 9 Science					
Paper	414	154	21		
Computer	466	181	25	-1.57	.12
Survey					
Paper	25.4	5.5	.77		
Computer	24.0	5.7	.80	1.22	.23
Comp-Writing					
Paper	9.9	2.6	.37		
Computer	9.1	3.0	.43	1.39	.17
WPM					
Paper	17.1	6.3	.88		
Computer	15.9	5.1	.72	1.01	.32

*Significant at the .05 level.

Note that statistical significance for the t-tests reported above was not adjusted to account for multiple comparisons. Given that six comparisons were made for each group, there is an increased probability that reported differences occurred by chance. Employing the Dunn approach to multiple comparisons (see Glass & Hopkins, 1984), α for *c* multiple comparisons, α_{pc}, is related to simple α for a single comparison as follows:

$$\alpha_{pc} = 1 - (1 - \alpha)^{1/c}$$

Hence, for six comparisons the adjusted value of a simple 0.05 alpha level becomes 0.009. Analogously, a simple alpha level of 0.01 for a simple comparison becomes 0.001.

Once the level of significance is adjusted for multiple comparisons, the open-ended science test is the only instrument for which there is a statistically significant group difference. This difference represents an effect size of .57 (Glass's delta effect size was employed). Although this effect size is about half of that reported by Russell and Haney (1997), it suggests that while half of the students in the computer group scored above 10.76, approximately 30% of students performing the test on paper scored above 10.76. The difference between the two groups' open-ended science scores, however, may be due in part to differences in their prior achievement as measured by SAT 9 Science scores.

To control for differences in prior achievement, a multiple regression was performed for each open-ended test. Tables 14 through 17 present the results of each test score regressed on the corresponding SAT 9 score and group membership. For all four regression analyses, the regression coefficient (B) for group membership indicates the effect group membership has on students' performance when the effect of SAT 9 scores is controlled. Group membership was coded 0 for the paper group and 1 for the computer group. A positive regression coefficient indicates that performing the test on computer has a positive effect on students' test performance. A negative regression coefficient suggests that on average students who performed the test on computer scored lower than students who performed the test on paper.

Table 14 indicates that SAT 9 Reading scores are a significant predictor of students' scores on the first open-ended language arts test. For each one standard score unit increase in SAT 9 Reading scores, on average students experience a .42 standard score increase in their test score. Table 14 also indicates that after controlling for differences in SAT 9 Reading scores, performing the first language arts test on computer has a negative impact on students scores. This effect, however, is not statistically significant.*

Table 14
Language Arts 1 Regressed on SAT 9 Reading and Group Membership

	B	SE B	Beta	T	Signif.
SAT 9 Reading	.007	.001	.42	4.87	<.0001
Group	-.443	.492	-.08	-.90	.37
F	11.85				<.0001
N	117				
R^2	.17				
Adjusted R^2	.16				

The results for the second language arts test are similar to those for the first language arts test. Table 15 shows that a one point standard score increase in SAT 9 Reading score is associated with a .4 point standard score increase in language arts 2 score and that this effect is statistically significant. Controlling for SAT 9 Reading scores, group membership does not have a significant effect on students' test score.

Table 15
Language Arts 2 Regressed on SAT 9 Reading and Group Membership

	B	SE B	Beta	T	Signif.
SAT 9 Reading	.005	.001	.40	4.3	<.0001
Group	.051	.428	.01	.1	.91
F	9.12				.0002
N	100				
R^2	.16				
Adjusted R^2	.14				

For both the math and science tests, SAT 9 scores and group membership have statistically significant effects on students' scores. The direction of the effect, however, is different for each test. Table 16 indicates that performing the open-ended math test on computer has a negative effect on students' test scores when SAT 9 Math scores are controlled. For science, this effect is reversed. Table 17 shows that after controlling for differences in SAT 9 Science scores, performing the open-ended science test on computer leads to higher scores than performing the same test on paper. For both tests, the effects are equivalent to just less than .2 standard score units.

Table 16
Math Regressed on SAT 9 Math and Group Membership

	B	SE B	Beta	T	Signif.
SAT 9 Math	.016	.001	.72	11.03	<.0001
Group	-1.546	.541	-.19	-2.86	.005
F	64.41				<.0001
N	110				
R^2	.54				
Adjusted R^2	.54				

Table 17
Science Regressed on SAT 9 Science and Group Membership

	B	SE B	Beta	T	Signif.
SAT 9 Science	0.014	.002	.59	7.50	<.0001
Group	1.466	.645	.18	2.28	.025
F	34.19				<.0001
N	102				
R^2	.41				
Adjusted R^2	.40				

Sub-Group Analyses

The regression analyses presented above indicate that mode of administration did not have a significant effect on students' performance on either language arts test. For the science test, performing the test on computer had a positive effect on students' scores. And for the math test, performing the test on computer led to lower performance. For all four of these analyses, the effect was examined across levels of computer use. To test whether the effect of mode of administration varied for students with different levels of computer skill, students' WPM was used to form three groups. The first group contained students whose WPM was .5 standard deviations below the mean, or less than 13.8. The second group contained students whose WPM was between .5 standard deviations below the mean and .5 standard deviations above the mean, or between 13.8 and 19.2. The third group contained students whose WPM was .5 standard deviations above the mean or greater than 19.2. For each group, the open-ended test scores were regressed on SAT 9 scores and group membership.

Table 18 displays the results of the three separate regressions for the first language arts test. For students whose WPM is .5 standard deviations below the mean and for students whose WPM is within .5 standard deviations of the mean, performing the test on computer has a negative effect on their scores. However, for these two groups of students, neither SAT 9 Reading nor group membership is a statistically significant predictor of language arts 1 score. However, for students whose keyboarding speed is one-half of standard deviation above the mean, or greater than 19.2 words per minute, performing the test on computer has a statistically significant positive effect on their performance. This effect is also three times stronger than the relationship between their SAT 9 reading score and their performance on the first language arts test. For the first language arts test, performing the test on computer seems to hurt students whose WPM is near or well below the mean and helps students whose WPM is well above the mean.

Table 18
Language Arts 1 Regressed on SAT 9 Reading and Group for Three Sub-Groups

WPM <13.8 N=30					
	B	SE B	Beta	T	Signif.
SAT 9 Reading	0.006	0.003	.36	1.98	.06
Group	-1.115	1.001	-.20	-1.12	.27
Adjusted R^2	.08				
13.8<WPM<19.2 N=54					
	B	SE B	Beta	T	Signif.
SAT 9 Reading	0.004	.002	.23	1.66	.10
Group	-1.330	.694	-.27	-1.91	.06
Adjusted R^2	.05				
WPM >19.2 N=33					
	B	SE B	Beta	T	Signif.
SAT 9 Reading	.003	.002	.19	1.32	.20
Group	2.946	.764	.56	3.86	.0006
Adjusted R^2	.38				

The same relationship was found for the second language arts test

(Table 19). However, for this test, the negative effect of taking the test on computer was statistically significant for students whose WPM was .5 standard deviations below the mean. For students whose WPM was within .5 standard deviations of the mean, performing the test on computer also had a negative effect on test performance, but this effect was not statistically significant. For students whose WPM was .5 standard deviations above the mean, performing the test on computer had a positive effect of nearly a half standard score on their language arts 2 test scores. This effect was statistically significant.

Table 19
Language Arts 2 Regressed on SAT 9 Reading and Group for Three Sub-Groups

WPM <13.8 N=35					
	B	SE B	Beta	T	Signif.
SAT 9 Reading	-.0001	.002	-.02	-.10	.92
Group	-1.48	.601	-.41	-2.47	.02
Adjusted R^2	.11				
13.8<WPM >19.2 N=37					
	B	SE B	Beta	T	Signif.
SAT 9 Reading	.002	.002	.17	1.02	.31
Group	-.974	.631	-.26	-1.54	.13
Adjusted R^2	.08				
WPM >19.2 N=28					
	B	SE B	Beta	T	Signif.
SAT 9 Reading	.006	.002	.37	2.32	.03
Group	2.068	.71	.46	2.90	.008
Adjusted R^2	.43				

For the open-ended math test, performing the test on computer had a negative effect on students' scores at all levels of keyboarding (Table 20). However, as keyboarding speed increased, this effect became less pronounced. For students whose WPM was .5 standard deviations below the mean, taking the test on computer had an effect of -.39 standard score units. For students whose WPM was within .5 standard deviations of the mean, this effect was -.24 standard score units. Both of these effects were statistically significant. However, for students whose WPM was .5 standard deviations above the mean, the effect was -.17 standard units and was not statistically significant.

Table 20
Math Regressed on SAT 9 Math and Group for Three Sub-Groups

WPM <13.8 N=30					
	B	SE B	Beta	T	Signif.
SAT 9 Math	.015	.003	.68	5.58	<.0001
Group	-3.068	.970	-.39	-3.16	.004
Adjusted R^2	.57				
13.8<WPM >19.2 N=49					
	B	SE B	Beta	T	Signif.
SAT 9 Math	.010	.003	.50	4.09	.0002
Group	-1.55	.796	-.24	-1.96	.05
Adjusted R^2	.30				
WPM >19.2 N=31					
	B	SE B	Beta	T	Signif.
SAT 9 Math	.019	.003	.73	5.96	<.0001
Group	-1.499	1.070	-.17	-1.40	.17
Adjusted R^2	.56				

Conversely, taking the science test on computer had a positive effect on students' scores at all levels of keyboarding speed (Table 21). However, this effect was only statistically significant for students whose WPM was within .5 standard deviations of the mean. For students whose WPM was .5 standard deviation units above the mean, this effect is less pronounced and is not statistically significant.

Table 21
Science Regressed on SAT 9 Science and Group for Three Sub-Groups

WPM <13.8 N=35					
	B	SE B	Beta	T	Signif.
SAT 9 Science	.011	.003	.50	3.33	.002
Group	.909	1.020	.13	.89	.38
Adjusted R^2	.24				
13.8<WPM >19.2 N=40					
	B	SE B	Beta	T	Signif.
SAT 9 Science	.010	.002	.47	4.00	.0003
Group	3.368	.893	.45	3.77	.0006
Adjusted R^2	.48				
WPM >19.2 N=27					
	B	SE B	Beta	T	Signif.
SAT 9 Science	.021	.004	.70	4.71	.0001
Group	.170	1.204	.02	.14	.89
Adjusted R^2	.45				

Discussion

The experiment described here extends the work of Russell and Haney (1997) and improved upon their study in five ways. First, this study included students whose prior computer experience varied more broadly. Second, many more open-ended items in the area of language arts, math and science were administered. Third, all of the open-ended test items included in this study had been used in state or national testing programs and had been validated previously. Fourth, an indicator of academic achievement was collected prior to the study and was used both to randomly assign students to groups and as a covariate during regression analyses. And fifth, information on students' prior computer use and keyboarding speed was collected and used during analyses.

In their study, Russell and Haney (1997) reported large, positive group differences which were consistent for all writing, math and science open-ended items administered on computer. In this study, a significant positive group difference was found only for the open-ended science test. This effect was about half the size reported by Russell and Haney (1997). However, in this study students' level of prior computer use varied more than it did in the previous study.

Although Russell and Haney did not collect a formal measure of computer use, the students included in their study were so accustomed to working on computer that when standardized tests were given, the school had difficulty finding enough pencils for all students. Although three years have passed since the previous study, it may be possible to estimate the difference in the level of prior computer use of the students included in both studies.

This study includes students from two schools, one of which was the focus of the previous study. Table 22 compares the WPM and survey scores for students in the ALL School and Sullivan Middle School. For both measures of computer use and for keyboarding speed, students in the ALL School have significantly higher scores. For the ALL School the mean WPM was nearly .5 standard deviations above the mean for the total sample while the mean for students from the Sullivan Middle School was below the total sample mean. By including Sullivan Middle School students in this study, a broader range and lower levels of computer use were represented. Including students with low levels of computer use and poor keyboarding skills seems to have counteracted the effect described in the previous study since these students performed less well on the language arts computer tests than on the paper tests.

Table 22
Comparison of Computer Use across Participating Schools

ALL N = 35 Sullivan N = 194	Mean	Std Dev	SE of Mean	t-value	Sig.
WPM					
ALL	18.9	5.1	.36		
Sullivan	16.1	6.2	1.05	2.94	.004
Survey					
ALL	27.5	5.8	.99		
Sullivan	23.8	5.5	.39	3.69	<.0001
Comp-Writing					
ALL	11.1	2.6	.43		
Sullivan	9.2	2.7	.20	3.84	<.0001

To examine the effect the mode of administration had on student performance at different levels of computer use, sub-group analyses were performed. Figure 1 summarizes the effects found for three sub-groups: a. students whose WPM was .5 standard deviations below the mean; b. students whose WPM was within .5 standard deviations of the mean; and c. students whose WPM was .5 standard deviations above the mean.

Figure 1 Effect of Performing Test on Computer When Prior Achievement is Controlled

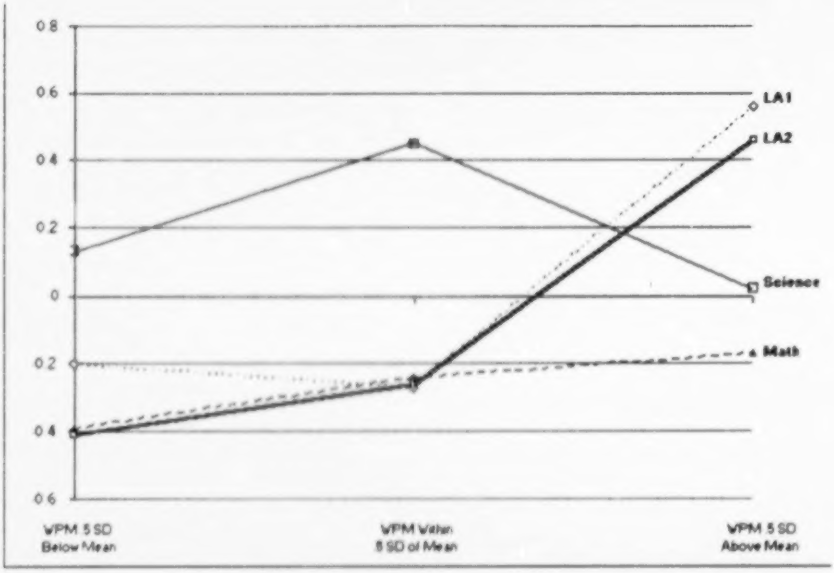


Figure 1 shows that across three of the four tests, performing the test on computer had an adverse effect on the performance of students whose WPM was .5 standard deviations below the mean. Conversely, for students whose WPM was at least .5 standard deviations above the mean, performing the language arts tests on computer had a moderate positive effect. While performing the math test on computer had a negative effect for all students, this negative effect became less pronounced as students' keyboarding speed increased. For the Science test, performing the test on computer had a positive effect across levels of computer use. However, the effect was much larger for students whose WPM was within .5 standard deviations of the mean than it was for students whose keyboarding speed was either .5 standard deviations above the mean or .5 standard deviations below the mean.

Explaining the Effects

To explore the reasons why some students had difficulty working on computers, students were asked to answer the following two questions after they completed the computer version of the test: 1. Do you think you would have done better on this test if you took it on paper? Why?; and 2. Besides not knowing the answer to a question, what problems did you have while taking this test on computer?

Students' responses to these questions were coded in two ways. First, the following numerical code was used:

- 0 - No, I would not perform better on paper
- 1 - I would perform the same or it didn't matter
- 2 - Yes, I would perform better on paper

In addition to these codes, an emergent coding scheme was used to tabulate the reasons students provided for their answers. While coding responses to the post-test questions, it became apparent that when read together, the two questions provided more information about students' experience than reading them separately. Some students would simply write yes or no for the first question, but their reasoning became apparent in their response to the second question. Other students explained the problems they encountered for the first question and wrote little for the second question. For this reason, responses to both questions were read during the emergent coding.

Table 23 presents the numerical codings for the first question. Across all tests, only 10% of students indicated that they would have performed better if they had taken the test on paper. However, over half of those who indicated they would have performed better on paper took the math test on computer. To explore why more students who took the math test on computer felt they would perform better on paper, the full responses to the two follow-up questions were examined.

Table 23
Frequency of Students Responses to Post-Test Question 1: Do you think you would have done better on this test if you took it on paper?

	Frequency	Percent
Language Arts 1		
Not better on paper	38	63.3
Same on paper	19	31.7
Better on paper	3	5.0
Language Arts 2		
Not better on paper	36	65.5
Same on paper	15	27.3
Better on paper	4	7.3
Math		
Not better on paper	20	35.7
Same on paper	18	32.1
Better on paper	18	32.1
Science		
Not better on paper	32	62.7
Same on paper	16	31.4
Better on paper	3	5.9

Table 24 presents the frequency of student responses by test. Clearly, the most frequently sited problem related to students' keyboarding skills.* Across all tests, about 25% of the students who performed the test on computer indicated they had difficulty "finding the keys," "pressing the wrong key" or simply said they "couldn't type." Twenty percent of the students who performed the math test on computer also complained that it was difficult to show their work on the computer or that they had to solve problems on paper and then transfer it to the computer. Several students who performed the language arts tests on computer mentioned that they preferred the computer because it was neater and that they didn't have to erase mistakes but could simply delete them. Across all tests, a few students also stated that they preferred the computer because their hands did not get as tired or that it was faster to write on the computer.

Table 24
Frequency of Responses to Post-test Questions 1 and 2

	LA 1	LA 2	Math	Science	Total
Difficulty typing	12	12	17	18	59
Neater on computer/can delete	8	9	2	2	21
Can't show work/drawings			10		10
Ran out of time on computer	1	2	3	1	7
Hand doesn't get tired	3	1	2	1	7
Faster on computer	3	1	2	1	7
Can take notes/solve problems on paper	1		4		5
Think better on computer/concentrate better	2	1	1	1	5
Write easier on paper			1	2	3
Hard looking back and forth between paper and computer			2		2
Hard to read screen			2		2
Problems with mouse		1	1		2
Easier to concentrate on paper	1	1			2
Write Poorly on paper				1	1
More comfortable on paper		1			1
More space on paper		1			1
Became confused where to put answers			1		1

Examining these responses, it appears that many more students who took the language arts tests recognized the computer's ability to display text that is easy to read and edit as an advantage. Conversely, students who took the math test felt that the inability to present and manipulate numbers in text was a disadvantage. In part, these different reactions to performing the tests on computer may explain the negative group effect for the math test and the positive group effects for the language arts tests. However, students' responses provide little insight into the overall group effect for science.

The Effect of WPM on Student Performance

To further examine the relationship between level of computer use and students' performance on the language arts tests, separate regression analyses were performed for students who performed the tests on paper and those who performed the tests on computer. For each of these regression analyses, the effect of prior computer use on students' performance was estimated controlling for SAT 9 scores. To provide separate estimates for keyboarding speed and for students' survey scores, two sets of regressions were performed for each sub- group. First, the test score was regressed on SAT 9 score, WPM and Survey. Second, the test score was regressed on SAT 9 score, WPM and Comp-Writing. Since Survey is partially composed of Comp-Writing, effects for each variable are estimated through separate regressions to avoid redundancy in the data and hence decrease the effects of colinearity. Figures 2 through 5 display the effects each variable had on test performance for students who took the test on computer and for those who took the test on paper.

Figure 2 and 3 show that across all tests, WPM is a weak predictor of students' scores when the test is performed on paper. However, for both language arts tests and the science test, WPM is a good predictor of students' scores when the test is performed on computer. This suggests that when these tests are performed on computer, the speed with which a student can type had a significant effect on their performance. However, for the math test, the effect of WPM on students' performance on computer is much less pronounced.

Figure 2: Effect of WPM on Student Performance Controlling for Survey

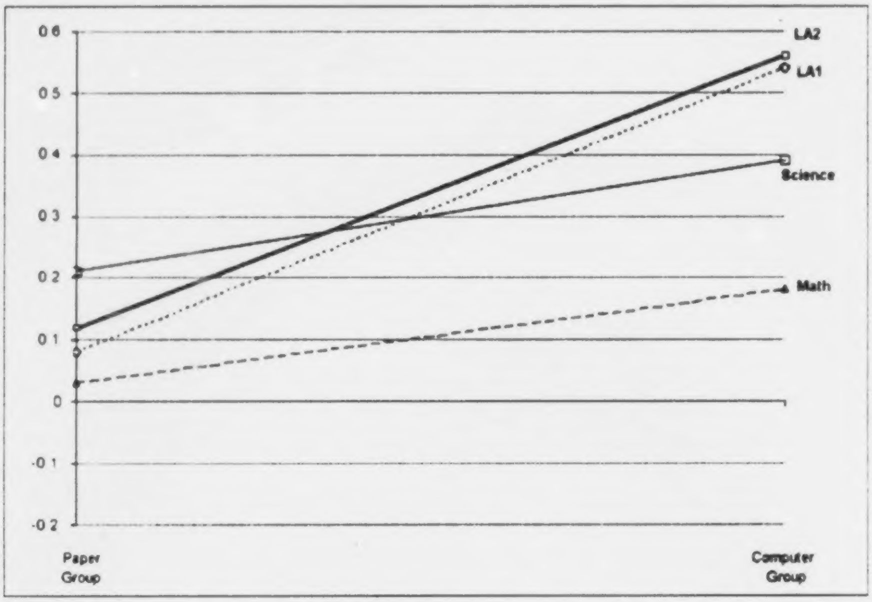
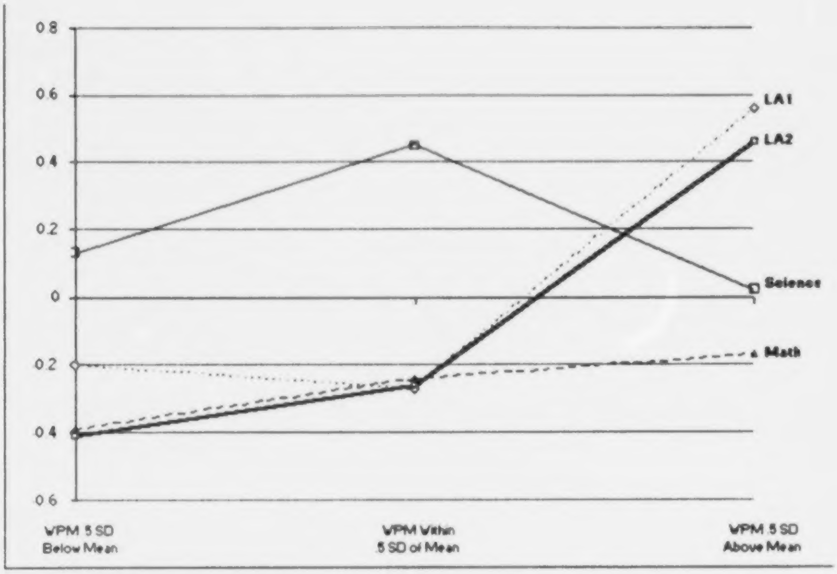


Figure 3: Effect of WPM on Student Performance Controlling for Comp-Writing



Figures 4 and 5 indicate that neither the total Survey score nor the Comp- Writing score had a meaningful effect on the performance of students in either group. In fact, when the effect of WPM is considered, both the amount of prior computer use and use of computers during the writing process have slightly lower effects when the test is taken on computer for the language arts and science tests. Yet, for the math test, the effect is larger and positive. This pattern is difficult to explain. Nonetheless, the weak relationship between either Survey or Comp-Writing and students' performance on computer suggests that students' level of computer use is not as important as their keyboarding proficiency in predicting their performance on open-ended tests. In future studies it is highly recommended that measures of keyboarding speed rather than self-reported levels of computer use are collected and used to examine effects of computer and paper administration.

Figure 4: Effect of Survey on Student Performance Controlling for WPM

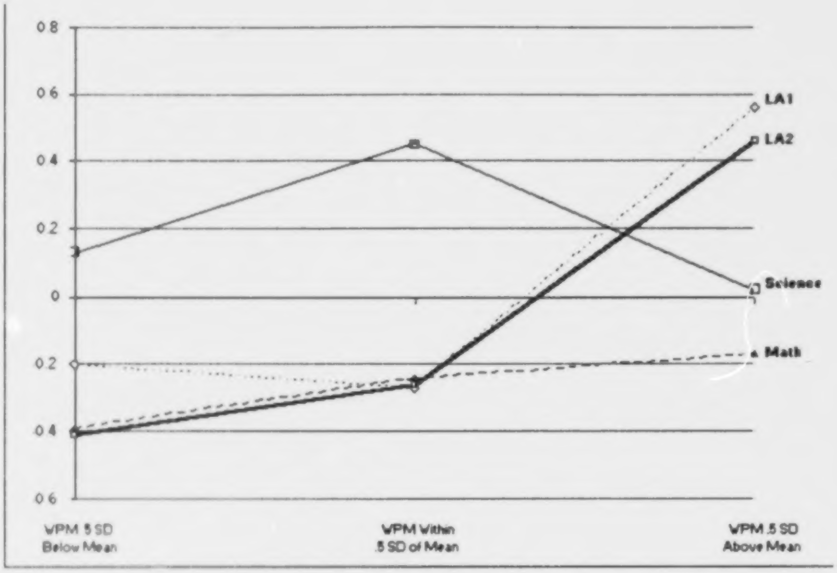
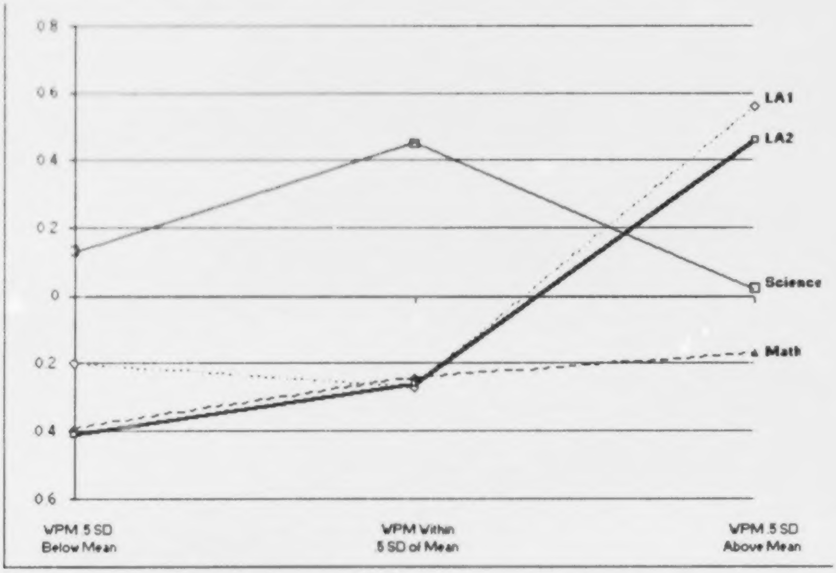


Figure 5: Effect of Comp-Writing on Student Performance Controlling for WPM



Preference and Performance

One of the questions this experiment was designed to address was whether students who performed the test via their preferred medium performed better than predicted and whether those who did not perform the test on their preferred medium performed worse than predicted. Prior to performing either test, students responded to the following survey question: If forced to choose, would you rather write a paper on computer or on paper? To examine the relationship between preference and performance, a dummy variable was coded 1 if the students' preference was the same as the medium on which they performed the test and 0 if their preference and performance medium did not match. For each test, students' test scores were regressed on their SAT 9 scores and Match. Table 25 shows that for the science test, students who took the test on their preferred medium did perform significantly better after controlling for prior achievement. Matching preference with medium of performance did not have a significant effect for the other three tests.

Table 25
Test Score Regressed on SAT 9 and Match

Language Arts 1 Match=43 NoMatch=74					
	B	SE B	Beta	T	Signif.
SAT 9 Reading	.007	.001	.42	4.83	<.0001
Match (1=yes)	-.367	.510	-.06	.72	.47
Adjusted R^2	.15				
Language Arts 2 Match=53 NoMatch=47					
	B	SE B	Beta	T	Signif.
SAT 9 Reading	.005	.001	.40	4.27	<.0001
Match (1=yes)	-.522	.422	-.11	1.24	.22
Adjusted R^2	.15				
Math Match=63 NoMatch=47					
	B	SE B	Beta	T	Signif.
SAT 9 Math	.016	.002	.72	10.80	<.0001
Match (1=yes)	-.948	.561	-.11	-1.69	.09
Adjusted R^2	.52				
Science Match=51 NoMatch=51					
	B	SE B	Beta	T	Signif.
SAT 9 Science	.014	.002	.58	7.46	<.0001
Match (1=yes)	1.487	.646	.18	2.30	.02
Adjusted R^2	.40				

As discussed above, preference for some students seems to have been influenced by social desirablity. As a result, the relationship between preference and performance on the preferred medium may be poorly estimated. Simply giving students the alternative to perform open-ended test questions via their "preferred" medium may not reduce the effect of medium found in this study. Rather, before students are given the choice, it might be useful to explain the apparent relationship between keyboarding speed and performance.

Gender, Keyboarding and Performance on Computers

Resent research suggests that females do not use computers in school as frequently as males (ETS, 1998). If this research is accurate, it is possible that the keyboarding skill of females is less developed than males. Given the relationship between WPM and performance on computer, performing tests on computer may have an adverse impact on the scores for females.

To examine the relationship between gender and WPM, an independent samples t-test was performed using all 229 students included in the study. To examine whether there were gender differences on computer use and prior achievement, t-tests were also performed for Survey, Comp-Writing and the SAT 9 comprehensive NCE. Table 26 indicates that WPM was the only variable for which there was a gender difference. However, on average, it was males' keyboarding speed that was 3 words per minute slower than females. This represents an effect size of approximately .68. This difference, however, does not seem to be caused by less computer experience or less use of computers in the writing process since there were negligible differences for either Survey or Comp-Writing.

Table 26
Gender Differences for WPM, Survey, Comp-Writing and SAT 9 Comprehensive

Males=97 Females=132	Mean	Std. Dev.	SE	T-value	Signif.
WPM					
Males	14.8	4.4	.45		
Females	17.8	5.6	.49	4.41	<.001
Survey					
Males	24.2	5.5	.56		
Females	24.5	5.8	.51	.37	.72
Comp-Writing					
Males	9.4	3.0	.31		
Females	9.6	2.6	.22	.56	.58
SAT 9 Comp.					
Males	408	160	16.3		
Females	398	143	12.4	.48	.63

As described above, WPM was a significant predictor for students' performance on computer in all subject areas. But given that males were on average slower keyboarders, one might expect their scores in all tests to be lower when performed on computer. Table 27 shows that this was the case for all four tests but that the difference was only significant for the first language arts test.

Table 27
Gender Differences for Test Performance on Computers

	Mean	Std. Dev.	SE	T-value	Signif.
LA 1					
Males (26)	4.96	2.60	.51		
Females (34)	6.50	3.22	.55	1.99	.05
LA 2					
Males (24)	4.33	2.57	.52		
Females (31)	5.48	.273	.49	1.59	.12
Math					
Males (21)	8.24	4.38	.96		
Females (35)	9.86	3.51	.59	1.52	.13
Science					
Males (21)	10.48	4.62	1.01		
Females (30)	10.97	3.83	.70	.41	.68

Table 28 shows that gender differences were not found for any tests when prior achievement and WPM were controlled. In part, this finding suggests that although males included in this study tended to be slower keyboarders, they performed as well as females with similar keyboarding and SAT 9 scores. This finding provides further evidence that keyboarding skills play an important role in how well students, regardless of their sex, perform on computers.

Table 28
Test Score Regressed on SAT 9, WPM and Gender for Computer Groups Only

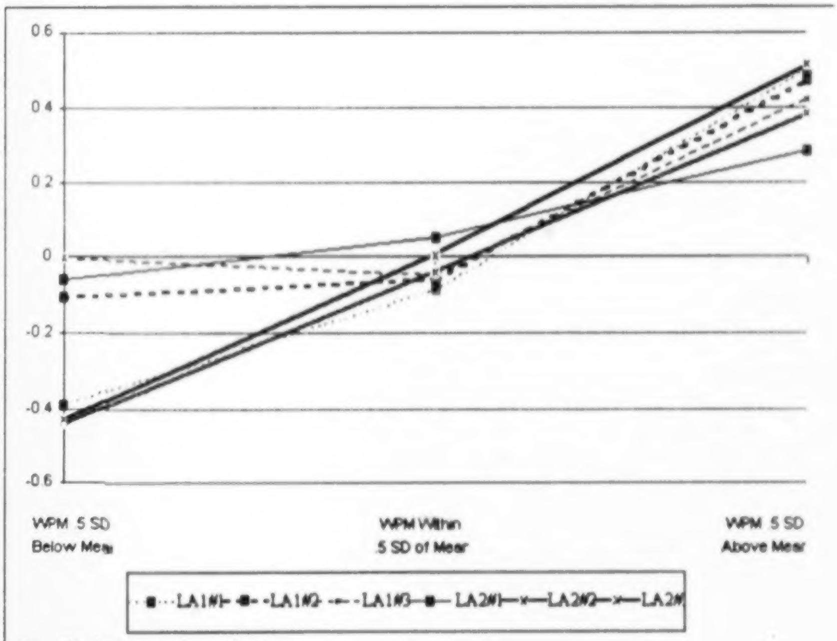
Language Arts 1 N=60					
	B	SE B	Beta	T	Signif.
SAT 9 Reading	.004	.002	.28	2.28	.03
WPM	.258	.072	.45	3.573	.0007
Sex (1=Male)	-.888	.648	-.15	1.37	.18
Adjusted R^2	.43				
Language Arts 2 N=55					
	B	SE B	Beta	T	Signif.
SAT 9 Reading	.003	.002	.20	1.51	.14
WPM	.246	.059	.59	4.15	.0001
Sex (1=Male)	.311	.605	.06	.51	.61
Adjusted R^2	.48				
Math N=56					
	B	SE B	Beta	T	Signif.
SAT 9 Math	.011	.002	.53	4.67	.0001
WPM	.177	.089	.23	2.01	.05
Sex (1=Male)	-.605	.833	-.08	.73	.47
Adjusted R^2	.45				
Science N=51					
	B	SE B	Beta	T	Signif.
SAT 9 Science	.011	.003	.49	4.25	.0001
WPM	.266	.095	.33	2.79	.008
Sex (1=Male)	-.379	.935	-.05	.41	.69
Adjusted R^2	.42				

Reading Comprehension vs. Writing Items

For students whose WPM was at least .5 standard deviations below the mean, performing either language arts test on computer had a negative effect on students' test scores. The negative effect was much larger for the second language arts test than it was for the first. To explore why the effect was larger for the second language arts test, the content of the two tests was examined.

Recall that the language arts tests contained two types of items, namely reading comprehension and writing. The first language arts test contained two reading comprehension items and only one writing item while the second language arts test contained two writing items and only one reading comprehension item. To examine the relationship between item type and the effect at each level of keyboarding speed, separate regressions were performed for each item. Figure 6 indicates that the effect for all items are about the same for students whose WPM is within .5 standard deviations of the mean or at least .5 standard deviations above the mean. However, for students whose WPM is at least .5 standard deviations below the mean, there seem to be two different effects. Language arts 1 item 1 and language arts 2 items 2 and 3 all seem to have an effect of about -.4. Language arts 1 item 2 and 3 and language arts 2 item 1 have effects between 0 and -.1. This pattern, however, does not seem to be related to item format. Although language arts 1 item 1 addresses reading comprehension, the two other items showing a similar effect test writing skills. Similarly, although language arts 1 item 2 and language arts 2 item 2 are both reading comprehension, the third item in this triad is a writing item. Thus, item format does not seem to explain the differences in the effect sizes for the two language arts tests at the low level of keyboarding speed.

Figure 6: Effects by Language Arts Item



Note: Black squares represent Reading
X's represent Writing items

Explaining Smaller Effect Sizes

As noted above, the magnitude of the effects in this study are about half the size reported by Russell and Haney (1997). While these positive effects are still quite large and represent approximately one half of a standard deviation difference in test scores, there are three observations that may shed some light on why the effects in this study were less pronounced than in the 1997 study.

First, the test scores used in Russell and Haney's study were part of a formal testing program. In the study reported here, the tests were described to students as practice for the spring MCAS administration and thus may not have been taken as seriously by students, especially those unaccustomed to working on computer. This was particularly evident during the computer administration. Whereas the author noted only one student being disciplined during four paper administration sessions that he observed, nearly 40 behavioral problems (e.g., students talking, students touching each other, or students moving around the room without permission) were addressed during the seven computer administration sessions that he observed. This increased level of disruptions may have occurred in part because students were frustrated by their inability to type. These disruptions also may have distracted students who did not experience difficulty keyboarding. Between not being as motivated during a practice test and being distracted more often, students' performance on computer may have suffered. In turn, this may have led to under-estimates of the positive effects and over-estimates of the negative effects.

Second, when the previous study occurred, the ALL school was in its third year of reform and was receiving full external support for its technology reforms. For this reason, there was great enthusiasm for the use of technology by teachers and students. As noted in the previous study, at the time students performed almost all of their work on computer. Since then, three years have passed, there has been a turn-over in teachers, and the external support for the ALL school has largely disappeared. For these reasons, it is possible that students in the ALL School are not using technology as extensively as they did three years ago. This is supported to some extent by the relatively low mean keyboarding speed for students in the ALL school. Although the ALL School's mean WPM was significantly higher than that of the Sullivan Middle School, it was still less than 20 WPM. This low keyboarding speed suggests that although students in the ALL School use computers more often than students in the Sullivan Middle School, their keyboarding skills are not as developed as one might expect if students are using computers on a daily basis. Although there is no direct data to confirm possible decrease use of computers in the ALL School, a decreased use might partially account for the smaller effect.

Finally, given the findings of the previous study and the heavy emphasis the Massachusetts Department of Education has placed on schools' performance on MCAS, it is possible that teachers require students to write more on paper in the ALL School now than three years ago in order to improve their performance on open-ended items. Sadly, after sharing this hypothesis with the ALL School's principal, Carol Shilinsky confirmed that in preparation for MCAS, teachers now require students to perform most of their writing on paper. If this was a successful strategy, then it would have improved students' scores on paper. In turn, the size of the effect of performing the tests on computer would be decreased.

Limitations

Despite efforts to create equivalent groups and to control for the confounding effects of scoring handwritten and computer printed responses, reading extensive passages of text on screen, and only using items that had been formally validated, this study still had several limitations. First, only a small group of students from one urban district were included. Recent research suggests that computers are not used the same way in all schools and that there are meaningful differences in the way students in urban and suburban schools use computers, particularly for math (ETS, 1998). These differences may lead to different effects for students in different settings.

Second, the tests were not administered under formal, controlled testing conditions. This may have decreased motivation, increased distractions and led to under-performance for many students. As noted above, this may be particularly true for students with better keyboarding skills who performed the tests on computer.

Third, although this study included many more open-ended items than did the previous study, testing time for each test was limited to sixty minutes. In order to increase the number of items included in the study, the time required to respond to items was limited, on average, to 10 minutes for math and science and 20 minutes for language arts. This time limit precluded extended writing and extended math items (requiring more than 10 minutes) from the study. However, MCAS and other testing programs include more extended open-ended items. And the effect of performing these types of items on computer may be larger given that in order to perform well, students generally need to produce more text.

Fourth, the sample of students included in this study had relatively slow keyboarding skills. For this reason, it was not possible to estimate the effect of taking open-ended tests on computer for students who are proficient or advanced keyboarders. Given the sharp increase in the size of the effect as keyboarding speed increases from near the mean to .5 standard deviations above the mean, it is possible that the effect of performing tests on computer is even larger for students with more advanced keyboarding skills.

Implications

This study suggests that for students who keyboard about 20 words per minute or more, performing open-ended language arts tests on paper substantially underestimates their level of achievement. However, for slower keyboarders, performing open-ended tests on computer adversely affects their performance. To provide more accurate estimates of students' achievement, these findings suggest that students who can keyboard at a moderate level should be allowed to compose their responses to open-ended items on computers. Conversely, students with weak keyboarding speed should compose their responses on paper.

This study also demonstrates that for math tests, performance on computer underestimates students' achievement regardless of their level of keyboarding speed. This occurred despite efforts to include items that did not require students to draw pictures or graphs to receive credit. Nonetheless, about 20% of the students who performed the math test on computer indicated that they had difficulty showing their work and/or needed scrap paper to work out their solutions. For these reasons, it is likely that the negative effect found in this study underestimates the effect that would occur if a full range of open- ended math items were included.

This study also re-emphasizes the danger of making inferences about students or schools based solely on paper-and-pencil tests. Similarly, as the public investigates the impact computers have on student learning (Oppenheimer, 1997), caution should be taken when student learning is measured by tests containing open-ended items. As found in the previous study, scores on paper and pencil tests for students accustomed to working on computer may substantially under- estimate students' achievement. As computer use in schools and at home continues to increase rapidly, it is likely that more students will develop solid keyboarding skills and, thus, will be adversely affected by taking open-ended tests on paper.

Finally, this study provides further evidence that the validity of open-ended tests should be considered in terms of both content and medium of learning. Until all students have access to and use computers regularly, open-ended tests administered via a single medium, either paper or computer, will likely under- estimate performance of students accustomed to working in the alternate medium. Based on this study, further research on a larger scale into computers and open- ended tests is clearly warranted. Until then, we should exercise caution when drawing inferences about students based on open-ended test scores when the medium of assessment does not match their medium of learning.

References

Bangert-Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research*, 63(1), 69-93.

Beaton, A. E. & Zwick, R. (1990). *The Effect of Changes in the National Assessment: Disentangling the NAEP 1985-86 Reading Anomaly*. Princeton, NJ: Educational Testing Service, ETS.

Bunderson, C. V., Inouye, D. K. & Olsen, J. B. (1989). The four generations of computerized educational measurement. In Linn, R. L., *Educational Measurement* (3rd Ed.), Washington, D.C.: American Council on Education, pp. 367-407.

Burstein, J., Kaplan, R., Wolff, S., & Lu, C. (1997). *Using lexical semantic techniques to classify free-responses*. A report issued by Educational Testing Service. Princeton, NJ.

Cizek, G. J. (1991). The effect of altering the position of options in a multiple-choice Examination. Paper presented at NCME, April 1991. (ERIC)

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Orlando, FL: Harcourt Brace Jovanovich College Publishers.

Daiute, C. (1986). Physical and cognitive factors in revising: insights from studies with computers. *Research in the Teaching of English*, 20 (May), p. 141-59.

Educational Testing Service. (1998). *Does it compute? The relationship between educational technology and student achievement in mathematics*. Princeton, NJ: Policy Information Center Research Division Educational Testing Service.

Etchison, C. (1989). Word processing: A helpful tool for basic writers. *Computers and Composition*, 6(2), 33-43.

Glass, G. & Hopkins, K. (1984). *Statistical Methods in Education and Psychology*. Boston, MA: Allyn and Bacon.

Glennan, T. K., & Melmed, A. (1996). *Fostering the use of educational technology: Elements of a national strategy*. Santa Monica, CA: RAND.

Grejda, G. F. (1992). Effects of word processing on sixth graders' holistic writing and revision. *Journal of Educational Research*, 85(3), 144-149.

Haas, C. & Hayes, J. R. (1986b). What did I just say? Reading problems in writing with the machine. *Research in the Teaching of English*, 20(1), 22- 35.

Haas, C. & Hayes, J. R. (1986a). Pen and paper versus the machine: Writers composing in hard-copy and computer conditions (CDC Technical Report No. 16). Pittsburgh, PA: Carnegie-Mellon University, Communication Design Center. (ERIC ED 265 563).

Hannafin, M. J. & Dalton, D. W. (1987). The effects of word processing on written composition. *The Journal of Educational Research*, 80 (July/Aug.) p. 338-42.

Hawisher, G. E. & Fortune, R. (1989). Word processing and the basic writer. *Collegiate Microcomputer*, 7(3), 275-287.

Hawisher, G. E. (1986, April). The effect of word processing on the revision strategies of college students. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC ED 268 546)

Kerchner, L. B. & Kistinger, B. J. (1984). Language processing/word processing: Written expression, computers, and learning disabled students. *Learning Disability Quarterly*, 7(4), 329-335.

Kurth, R. J. (1987). Using word processing to enhance revision strategies during student writing activities. *Educational Technology*, 27(1), 13-19.

MacArthur C. & Graham, S. (1987). Learning disabled students' composing under three methods of text production: Handwriting, word processing and dictation. *Journal of Special Education*, 21(3), 22-42.

MacArthur, C. A. (1988). The impact of computers on the writing process. *Exceptional Children*, 54(6), 536-542.

Mourant, R. R., Lakshmanan, R. & Chantadisai, R. (1981). Visual Fatigue and Cathode Ray Rube Display Terminals. *Human Factors*, 23(5), 529-540.

Nichols, L. M. (1996). Pencil and paper versus word processing: a comparative study of creative writing in the elementary school. *Journal of Research on Computing in Education*, 29(2), 159-166.

Oppenheimer, T. (1997). The computer delusion. *The Atlantic Monthly*. 280(1), 45-62.

Owston, R. D. (1991). Effects of word processing on student writing in a high-computer-access environment (Technical Report 91-3). North York, Ontario: York University, Centre for the Study of Computers in Education.

Phoenix, J. & Hannan, E. (1984). Word processing in the grade 1 classroom. *Language Arts*, 61(8), 804-812.

Powers, D., Fowles, M, Farnum, M, & Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement*, 31(3), 220-233.

Robinson-Stavely, K. & Cooper, J. (1990). The use of computers for writing: Effects on an English composition class. *Journal of Educational Computing Research*, 6(1), 41-48.

Russell, M. & Haney, W. (1997). Testing writing on computers: an experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives*, 5(3). Available online at http://epaa.asu.edu/epaa/v5n3.html.

Sitko, M.C. & Crealock, C. M. (1986, June). A longitudinal study of the efficacy of computer technology for improving the writing skills of mildly handicapped adolescents. Paper presented at at the Invitational Research Symposium on Special Education Technology, Washington, DC.

Snyder, T. D. & Hoffman, C. (1990). *Digest of Education Statistics*. Washington, DC: U. W. Department of Education.

Snyder, T. D. & Hoffman, C. (1994). *Digest of Education Statistics*. Washington, DC: U. W. Department of Education.

Vacc, N. N. (1987). Word processor versus handwriting: A comparative study of writing samples produced by mildly mentally handicapped students. *Exceptional Children*, 54(2), 156-165.

Williamson, M. L. & Pence, P. (1989). Word processing and student writers. In B. K. Briten & S. M. Glynn (Eds.), *Computer Writing Environments: Theory, Research, and Design* (pp. 96-127). Hillsdale, NJ: Lawrence Erlbaum & Associates.

Zandvliet, D. & Farragher, P. (1997). A comparison of computer-administered and written tests. *Journal of Research on Computing in Education*, 29(4), 423-438.

About the Author

Michael Russell
Email: russelmh@bc.edu

Michael Russell is a research fellow for the National Board on Educational Testing and Public Policy and a research associate in the Center for the Study of Testing, Evaluation and Educational Policy at Boston College. His research interests include standards based reform, assessment, and educational technology.

Copyright 1999 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ

85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Green Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
New York University

Dennis Sayers
Ann Leavenworth Center for Accelerated Learning

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

David D. Williams
Brigham Young University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina—Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary McKeown-Moak
MGT of America (Austin, TX)

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois—UC

Robert T. Stout
Arizona State University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)
Universidad de Guadalajara
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)
Universidad de Cádiz
felix.angulo@uca.es

Teresa Bracho (México)
Centro de Investigación y Docencia Económica-CIDE
bracho dis1.cide.mx

Alejandro Canales (México)
Universidad Nacional Autónoma de México
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)
Arizona State University
casanova@asu.edu

José Contreras Domingo
Universitat de Barcelona
Jose.Contreras@doe.d5.ub.es

Erwin Epstein (U.S.A.)
Loyola University of Chicago
Eepstein@luc.edu

Josué González (U.S.A.)
Arizona State University
josue@asu.edu

Rollin Kent (México)
Departamento de Investigación Educativa-DIE/CINVESTAV
rkent@gemtel.com.mx
kentr@data.net.mx

María Beatriz Luce (Brazil)
Universidad Federal de Rio Grande do Sul-UFRGS
lucemb@orion.ufrgs.br

Javier Mendoza Rojas (México)
Universidad Nacional Autónoma de México
javiermr@servidor.unam.mx

Marcela Mollis (Argentina)
Universidad de Buenos Aires
mmollis@filo.uba.ar

Humberto Muñoz García (México)
Universidad Nacional Autónoma de México
humberto@servidor.unam.mx

Angel Ignacio Pérez Gómez (Spain)
Universidad de Málaga
aiperez@uma.es

Daniel Schugurensky (Argentina-Canadá)
OISE/UT, Canada
dschugurensky@oise.utoronto.ca

Simon Schwartzman (Brazil)
Fundação Instituto Brasileiro e Geografia e Estatística
simon@openlink.com.br

Jurjo Torres Santomé (Spain)
Universidad de A Coruña
jurjo@udc.es

Carlos Alberto Torres (U.S.A.)
University of California, Los Angeles
torres@gseisucla.edu

---

archives | abstracts | editors | board | submit | comment | subscribe | search